Electronic Supplementary Material (ESI) for Chemical Communications.
This journal is © The Royal Society of Chemistry 2017

Chemoselective and site-selective peptide and native protein modification enabled by aldehyde auto-oxidation

Landa Purushottam, Srinivasa Rao Adusumalli, Maheshwerreddy Chilamari and Vishal Rai*

Department of Chemistry, Indian Institute of Science Education and Research (IISER) Bhopal, Bhopal, India.

vrai@iiserb.ac.in

Contents

1. General

The reagents, proteins, and enzymes were purchased from Sigma-Aldrich, Alfa Aeser and Merck Novabiochem. Histone (H 3.1) was expressed and purified using literature protocol.[1] The organic solvents used were reagent grade. Aqueous buffers were prepared freshly using Millipore Grade I water (Resistivity > 5 MΩ cm, Conductivity < 0.2 μS/cm, TOC <30 ppb) and the pH was adjusted using pH meter Mettler Toledo (FE20). Phosphate buffer (pH 7.0, 0.1 M) was prepared by mixing the appropriate volume of 0.2 M monobasic sodium phosphate solution (4.25 ml) and 0.2 M dibasic sodium phosphate solution (45.75 ml) and diluting to a total volume of 100 ml, to make 0.1 M phosphate buffer of required pH. The pH of the buffer was adjusted with dibasic or monobasic sodium phosphate solution if required. Tris buffer (pH 7.8, 0.01 M $CaCl_2$, 0.4 M) was prepared by dissolving tris(hydroxymethyl)aminomethane (2.42 g) and $CaCl_2$ (220 mg) in 50 ml Millipore Grade I water and the pH was adjusted to 7.8 by using 6N HCl. The reaction mixture was either stirred (Heidolph, 500-600 rpm), or vortexed in incubator-shaker Thermo Scientific MaxQ 8000 (350 rpm, 25-37 ^{o}C). Cellulose membrane (MWCO, 6-8 kD) from Spectrum labs was used for dialysis. BUCHI Rotavapor R-210/215 was used to remove organic solvents and CHRiST ALPHA 2-4 LD plus lyophilizer was used for lyophilization of aqueous samples. The peptide was synthesized by SPPS using Fmoc chemistry on Biotage Syro I parallel peptide synthesis system. Circular Dichroism (CD) measurements were recorded on JASCO J-815 CD spectropolarimeter equipped with peltier temperature controller. All the spectra were measured with a scan speed of 50 nm/min, spectral bandwidth 1 nm using 2 cm path length cuvette at 25 ^{o}C. In fluorescence spectroscopy, emission measurements were carried out in HORIBA JOBIN YVON, FLUOROLOG 3-111. The fluorescence spectra were measured with a quartz cuvette of 1 mm path length. The protein A280 application on Thermo Scientific NanoDrop^{TM} 2000/2000c spectrophotometer and Bradford assay was used for the estimation of protein concentration.

Chromatography: Thin-layer chromatography (TLC) was performed on silica gel coated aluminium TLC plates (Merck, TLC Silica gel 60 F_{254}) and visualized using a UV lamp (254 nm) and stains such as iodine, ninhydrin, 2,4-diphenylhydrazine. For reactions where chromatography was involved, Flash column chromatography was carried out on Combiflash Rf 200 or gravity columns using 230-400 mesh silica gel from Merck.

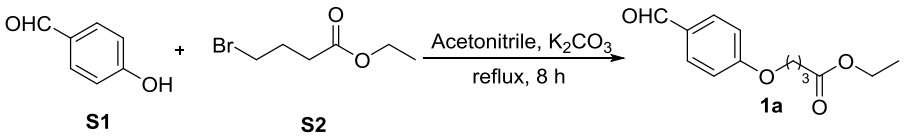
Nuclear magnetic resonance spectra: ^{1}H, ^{13}C, DQF-COSY and HSQC NMR spectra were recorded on Bruker Avance III 400, 500 and 700 MHz NMR spectrometer. ^{1}H NMR spectra were referenced to TMS (0 ppm), DMSO-d_6 (2.50 ppm) and MeOH-d_4 (3.31 ppm), ^{13}C NMR spectra were referenced to $CDCl_3$ (77.16 ppm), MeOH-d_4 (49.01 ppm) and DMSO-d_6 (39.52 ppm). Peak multiplicities are designated by the following abbreviations: s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; dd, doublet of doublet; dt, doublet of triplet; td, triplet of doublet, ddd, doublet of doublet of doublet. Spectra were recorded at 298 K.

Mass spectrometry: Agilent Technologies 1200 series HPLC paired to Agilent 6130 mass spectrometer (ESI/APCI) was used for LC-MS data. This was used to estimate conversion of peptides **6** to formylated peptides **7**. HPLC experiments of compounds **7a** and **7g** were performed on Zorbax 300SB-C18 StableBond Analytical column (4.6 × 250 mm × 5μ) with flow rate 0.4 ml/min. HPLC experiments of compounds **7b-7f**, **7h-j**, and peptides in additional discussion of supplementary information were performed on Poroshell 120 EC-C18 column (3.0 × 50 mm × 2.7 μ) with flow rate 0.4 ml/min. Bruker

Daltonics MicroTOF-Q-II with electron spray ionization (ESI) was used for HRMS data. Matrix-assisted laser desorption/ionization time of flight mass spectrometry was performed with Bruker Daltonics UltrafleXtreme Software-Flex control version 3.4, using sinapic acid and α-cyano-4-hydroxycinnamic acid (HCCA) matrix. Data analysis was performed using flex analysis. Peptide mass and fragment ion calculator (http://db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html) were used for peptide mapping and sequencing.

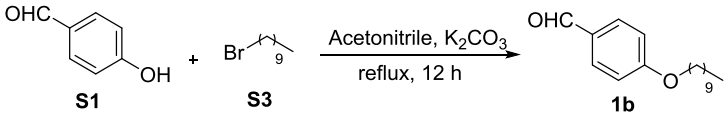
2. Procedures for synthesis of starting materials

Synthesis of ethyl 4-(4-formylphenoxy)butanoate (1a)[2]



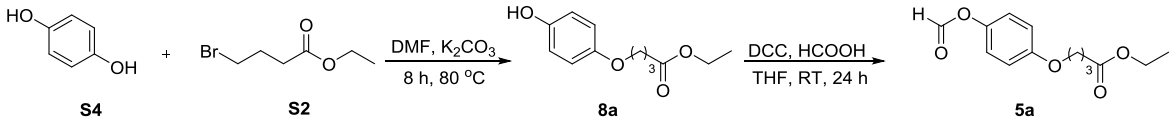
In a 25 ml round bottom flask, *p*-hydroxy benzaldehyde (122.1 mg, 1 mmol) was dissolved in 10 ml acetonitrile. To this aldehyde solution, K_2CO_3 (276.4 mg, 2 mmol) and ethyl 4-bromobutanoate (0.17 ml, 1.2 mmol) were added and the reaction mixture was allowed to reflux for 8 h. The reaction was monitored using thin layer chromatography and upon completion, the reaction mixture was filtered to remove potassium carbonate. The solution was concentrated under vacuum and the product was purified using flash column chromatography (ethyl acetate:*n*-hexane 5:95) to afford ethyl 4-(4-formylphenoxy)butanoate **1a** (93% yield, 220 mg). TLC (ethyl acetate:*n*-hexane 5:95, R_f 0.52)**,** ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.88 (s, 1H), 7.83 (d, *J* = 8.6 Hz, 2H), 6.99 (d, *J* = 8.7 Hz, 2H), 4.15 (q, *J* = 7.2 Hz, 2H), 4.11 (t, *J* = 6.2 Hz, 2H), 2.53 (t, *J* = 7.2 Hz, 2H), 2.15 (m, *J* = 6.7 Hz , 2H), 1.26 (t, *J* = 7.1 Hz, 3H) ppm. ^{13}C NMR (101 MHz, $CDCl_3$) δ 190.9, 173.1, 164.0, 132.1, 130.1, 114.9, 67.3, 60.7, 30.7, 24.5, 14.3 ppm. MS (ESI) $[MH]^+$ calcd. $C_{13}H_{16}O_4$ 237.1, found 237.1.

Synthesis of 4-(decyloxy)benzaldehyde (1b)



In a 25 ml round bottom flask, *p*-hydroxy benzaldehyde (122.1 mg, 1 mmol) was dissolved in 10 ml acetonitrile. To this solution, K_2CO_3 (276.4 mg, 2 mmol) and 1-bromodecane (0.265 ml, 1.2 mmol) were added and the reaction mixture was allowed to reflux. The reaction was monitored using thin layer chromatography. After completion (12 h), the reaction mixture was filtered to remove potassium carbonate. The filtrate was concentrated under reduced pressure and the product was purified using flash column chromatography (*n*-hexane) to afford ethyl 4-(decyloxy)benzaldehyde **1b** (85% yield, 223 mg). TLC (*n*-hexane, R_f 0.71)**,** ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.90 (s, 1H), 7.89- 7.81 (m, 2H), 7.06-6.96 (m, 2H), 4.06 (t, *J* = 6.6 Hz, 2H), 1.88-1.77 (m, 2H), 1.55-1.43 (m, 2H), 1.43-1.22 (m, 12H), 0.90 (q, *J* = 6.5 Hz, 3H) ppm. ^{13}C NMR (125 MHz, $CDCl_3$) δ 190.9, 164.4, 132.1, 129.8, 114.8, 68.5, 32.0, 29.6, 29.1, 26.0, 22.8, 14.2 ppm. MS (ESI) $[MH]^+$ calcd. $C_{17}H_{26}O_2$ 263.3, found 263.2.

Synthesis of 4-(4-formylphenoxy)butanoate (8a)

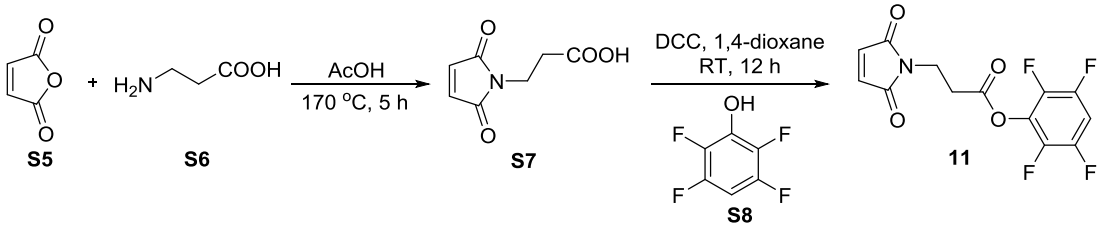


In a 25 ml round bottom flask, 4-hydroxyphenol (220 mg, 2 mmol) was dissolved in DMF (8 ml). To this solution, K_2CO_3 (276.4 mg, 2 mmol) and ethyl-4-bromobutanoate (0.286 ml, 2 mmol) were added. The resulting mixture was stirred at 80 ^oC for 8 h followed by filtration to remove potassium carbonate. The filtrate was extracted using ethyl acetate:*n*-hexane (4:6, 3 × 20 ml), organic layers were combined and washed with brine solution. The organic layer was dried over Na_2SO_4, concentrated and the residue was purified using flash column chromatography (ethyl acetate:*n*-hexane 10:90), to afford **8a** (31% yield, 140 mg). TLC (ethyl acetate:*n*-hexane 10:90, R_f 0.64), ^1H NMR (500 MHz, $CDCl_3$) δ 6.80 (m, 4H), 6.74-6.70 (m, 1H), 4.14 (q, *J* = 7.1 Hz, 2H), 3.94 (t, *J* = 6.1 Hz, 2H), 2.50 (t, *J* = 7.3 Hz, 2H), 2.13-2.03 (m, 2H), 1.26 (t, *J* = 7.1 Hz, 3H) ppm. ^{13}C NMR (125 MHz, $CDCl_3$) δ 173.7, 152.8, 149.8, 116.0, 115.6, 67.4, 60.6, 30.9, 24.7, 14.1ppm. MS (ESI) $[MH]^+$ calcd. $C_{12}H_{16}O_4$ 225.1, found 225.1.

Synthesis of ethyl 4-(4-(formyloxy)phenoxy)butanoate (5a)

A solution of N, N′-dicyclohexylcarbodiimide (185 mg, 0.9 mmol) in anhydrous THF (2.2 ml) was added dropwise to a stirred solution of 4-(4-formylphenoxy)butanoate (100 mg, 0.45 mmol) and formic acid (0.05 ml, 1.2 mmol) in an ice bath. After complete addition, the ice bath was removed and the reaction mixture was allowed to stir at ambient temperature. The reaction was monitored using thin layer chromatography. After complete consumption of **5** (24 h), the reaction mixture was filtered. The filtrate was concentrated under reduced pressure and subjected to flash column chromatography (ethyl acetate:*n*-hexane 5:95) to afford **5a** (45% yield, 50 mg). TLC (ethyl acetate:*n*-hexane 4:6, R_f 0.76), ^1H NMR (500 MHz, $CDCl_3$) δ 8.28 (s, 1H), 7.06-7.04 (m, 2H), 6.91-6.86 (m, 2H), 4.15 (q, *J* = 7.1 Hz, 2H), 4.00 (t, *J* = 6.1 Hz, 2H), 2.51 (t, *J* = 7.3 Hz, 2H), 2.14-2.07 (m, 2H), 1.30-1.22 (m, 3H) ppm. ^{13}C NMR (125 MHz, $CDCl_3$) δ 173.3, 159.8, 157.1, 143.5, 122.1, 115.4, 67.3, 60.6, 30.8, 24.7, 14.3 ppm. HRMS (ESI) $[MNa]^+$ calcd. $C_{13}H_{16}O_5$ 275.0895, found 275.0890.

Synthesis 3-(2, 5-dioxo-2, 5-dihydro-1H-pyrrol-1-yl)propanoic acid (S7)[3]



In a 25 ml round bottom flask, maleic anhydride (392 mg, 4 mmol) and β-alanine (427 mg, 4.8 mmol) were dissolved in glacial acetic acid (10 ml). The reaction mixture was allowed to reflux for 5 h and reaction was monitored using TLC. After completion of the reaction, acetic acid was removed on rotary evaporator as azeotropic mixture with 1,4-dioxane (4 ml). The reaction mixture was dissolved in ethylacetate (20 ml) followed by removal of a white precipitate (β-alanine). The filtrate was concentrated under reduced pressure and subjected to flash column chromatography (MeOH:DCM 5:95) to obtain the analytically pure product **S7** (55% yield, 372 mg). TLC (methanol:dichloromethane 5:95, R_f 0.48), ^1H

NMR (500 MHz, D_2O) δ 6.81 (t, *J* = 3.8 Hz, 2H), 3.81-3.69 (t, 2H), 2.62-2.57 (m, 1H) ppm. ^{13}C NMR (125 MHz, D_2O) δ 175.6, 172.1, 135.0, 34.5, 31.3 ppm. MS (ESI) $[MH]^+$ calcd. $C_7H_7NO_4$ 170.0, found 170.0.

2,3,5,6-tetrafluorophenyl 3-(2, 5-dioxo-2, 5-dihydro-1H-pyrrol-1-yl)propanoate (11)[4]

In a 10 ml round bottom flask, maleimido-propionic acid (50 mg, 0.29 mmol) and tetrafluorophenol (50 mg, 0.29 mmol) were dissolved in anhydrous dioxane (1.5 ml). DCC (89.6 mg, 0.43 mmol) in dioxane (1.5 ml) was added to this solution and the mixture was stirred at room temperature for 12 h. The solution was filtered and the filtrate was concentrated using rotary evaporator. The crude reaction mixture was dissolved in DCM and stored at -20 ^oC for 12 h. The precipitated dicyclohexylurea was filtered and the filtrate was concentrated under reduced pressure. The crude mixture was subjected to flash column chromatography (ethyl acetate:*n*-hexane 5:95) to obtain the pure product **11** (32% yield, 29.4 mg). TLC (ethyl acetate:*n*-hexane 5:95, R_f 0.48), ^1H NMR (400 MHz, $CDCl_3$) δ 6.98 (m, *J* = 9.9, 7.1 Hz, 1H), 6.73 (s, 2H), 3.96 (t, *J* = 7.1 Hz, 2H), 3.05 (t, *J* = 7.1 Hz, 2H) ppm. ^{13}C NMR (126 MHz, $CDCl_3$) δ 170.1, 166.6, 147.0, 145.0, 139.3, 134.3, 103.4, 33.2, 31.8 ppm. MS (ESI) $[MH]^+$ calcd. $C_{13}H_7F_4NO_4$ 318.0, found 318.0.

Peptide synthesis

Solid phase peptide synthesis (Biotage Syro I peptide synthesizer) with standard Fmoc-protecting group strategy was used for generating the required pool of peptides. Rink amide resin (loading capacity: 0.78 mmol/g) was used for synthesis of all the peptides. All Fmoc amino acids were activated by *in situ* HBTU/DIPEA activation procedure. The cleavage from the solid support and the simultaneous deprotection of all side chain residues were performed by suspending the fully protected compound resin in TFA:H_2O:TIS (95:2.5:2.5) for 3 h. The analytically pure peptides were isolated by precipitation in cold diethyl ether or by reverse phase preparative HPLC.

3. Procedures for the modification of peptides and proteins

Procedure for selective formylation of peptides

In a 5 ml vial, peptide **6** (0.01 mmol) was dissolved in phosphate buffer (1.8 ml, 0.1 M, pH 7.0). To this solution, freshly prepared chemically pure 4-(decyloxy)benzaldehyde **1b** (2 mmol) in DMF (0.2 ml) was added and vortexed (350 rpm) at 25 ^oC. The overall concentration of peptide and aldehyde was 5 mM and 1 M respectively. The progress of reaction was monitored by reversed-phase HPLC-ESI-MS. After 48 h, the reaction mixture was concentrated by lyophilization. Purification of the crude reaction mixture was performed by silica gel (230-400 mesh) flash column chromatography using methanol/chloroform (1:9) resulting in analytically pure products. For the peptides with excellent conversion (>99%), reaction mixture was triturated with diethyl ether (10 × 1 ml) to remove the unreacted 4-(decyloxy)benzaldehyde **1b** and co-products such as 4-(decyloxy)phenol **8b** and 4-(decyloxy)benzoic acid **4b**. Subsequently, desalting was performed by addition of methanol that extracts the product leaving the insoluble salts in vial. Concentration of the solution gave analytically pure products. All the purified samples were analyzed by NMR spectroscopy (^1H, ^{13}C, DQF-COSY and HSQC), LRMS and HRMS.

Note: Supply of additional oxygen by purging and stoichiometry of aldehyde above 600 equivalents led to a significant number of unidentified side products. A liquid-liquid extraction with ethylacetate and hexane (9:1) removes excess aldehyde **1b**, carboxylic acid **4b** and phenol derivative **8b**.

General procedure for site-selective formylation of proteins

In a 2 ml Eppendorf tube, protein (7.3 nmol) in phosphate buffer (95 µl, 0.1 M, pH 7.0) was taken. Aldehyde, ethyl-4-(4-formylphenoxy)butanoate **1a** (4.38 µmol) in DMSO (5 µl) was taken from a freshly prepared stock solution and added to the protein followed by vortexing (350 rpm) at 25 ^{o}C or 37 ^{o}C. The overall concentration of protein and aldehyde was 73 µM and 43.8 mM respectively. The progress of the transformation was followed by sampling aliquots at various time intervals by MALDI-ToF-MS using sinapic acid as matrix. After 24-48 h, the reaction mixture was further diluted with water (0.8 ml) and unreacted aldehyde was removed with spin filtration (3 kDa) (Sartorius VIVASPIN 500) and the sample was concentrated by lyophilization before subjecting it to digestion, peptide mapping and sequencing by MS-MS. In the case of Melittin, concentration of aldehyde (21.9 mM, 2.19 µmol) was reduced to control the selectivity. All the other steps were similar to other proteins. In the case of RNase A (7.3 nmol), 4-(decyloxy)benzaldehyde **1b** (4.38 µmol) was used as formylating pre-reagent.

Procedure for interception of formylation

In a 2 ml Eppendorf tube, RNase A (7.3 nmol) in phosphate buffer (95 µl, 0.1 M, pH 7.0) was taken. Aldehyde, ethyl 4-(4-formylphenoxy)butanoate **1a** (4.38 µmol) in DMSO (5 µl) was taken from a freshly prepared stock solution and added to the protein followed by vortexing (350 rpm) at 25 ^{o}C. After 3 h, 2,3,5,6-tetrafluorophenyl 3-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)propanoate (**11**, 109.5 nmol) in DMSO (3 µl) from freshly prepared stock solution was added to the reaction mixture gradually over a period of 75 minutes. The concentration of the protein, aldehyde and **11** in the reaction mixture was 73 µM, 43.8 mM and 1.09 mM respectively. The reaction was stopped in another 15 minutes by extraction of unreacted electrophile **11** and aldehyde **1a** using ethyl acetate/hexane (8:2, 5×1 ml). The reaction mixture was subjected to dialysis and the sample was concentrated by lyophilization. Mono-labeled protein 10m was dissolved in phosphate buffer (95 µl, 0.1 M, pH 7.0) and treated with 4-methyl-7-mercaptocoumarin (**12**, 3.65 mM, 365 nmole, from freshly prepared solution) in DMSO (5 µl) to give the fluorescent labeled protein **13**. Protein labeling was monitored by using MALDI-ToF. After the late-stage modification, unreacted **12** was extracted using ethyl acetate/hexane (8:2, 5×1 ml). The reaction mixture was subjected to dialysis and the sample was concentrated by lyophilization, before subjecting it to digestion and peptide mapping.

Procedure for in-solution digestion of RNase A, Lysozyme C and Histone (H 3.1)

All solutions were prepared freshly before use in reactions.[5]

Step 1. Protein (0.1 mg) in 100 mM tris (10 µl, pH 7.8) with urea (6 M) was taken in 2 ml Eppendorf tube and vortexed for 30 minutes.
Step 2. **Disulfide reduction:** To reduce the disulfide bonds, reducing agent (1 µL, 0.2 M DTT in 0.1 M tris pH 7.8) was added to the solution and sample was vortexed for 1 h at 25 °C or 37 °C.
Step 3. **Sulfhydryl alkylation:** To block the free sulfhydryl groups, alkylating agent (4 µL, 0.2 M iodoacetamide in 0.1 M tris pH 7.8) was added to the solution and incubated (in the dark) for 1 h at room temperature.
Step 4. **Quenching alkylating reagent:** To quench the unreacted iodoacetamide, reducing agent (4 µL) was added again to the mixture and the sample was vortexed at 25 °C for 1 h. Dilution of the reaction mixture with grade I water reduced the urea concentration to 0.6 M.
Step 5. **Enzymatic digestion:** To this solution, 10 µL of enzyme solution [2 µg, based on ratio of enzyme/protein (1:50); enzyme in 1 mM HCl was dissolved in 0.1 M tris pH 7.8 and 0.01 M $CaCl_2$] was added and the mixture was incubated at 37 °C for 18 h. Trifluoroacetic acid (0.5 %) was used to adjust the pH of digested solution to < 6 (verified by pH paper). Subsequently, the sample was used for peptide mapping by MS and sequencing by MS-MS investigations.

Procedure for in-solution digestion of Ubiquitin, Melittin, Cytochrome C and Myoglobin

Steps 1 and 5 were used for digestion of Ubiquitin, Melittin, Cytochrome C and Myoglobin. Steps 2, 3 and 4 are not desired as these proteins do not have disulfide bridges and derived free sulfhydryl groups.

α-Chymotrypsin is used for digestion of RNase A, Melittin, Ubiquitin, Lysozyme C, Myoglobin, Cytochrome C, whereas, trypsin is used for Histone (H 3.1).

4. Reference table: characterization data

Figure/Table	Compound	Related data	Page
Fig. 2	**1a**, **2c**, **4a**, **5a**, **8a**	Fig. S3, Fig. S4, Sec. 6	S3, S4, S10
Fig. 3	**7a-j**	Sec. 6	S20, S21, S22, S23, S24
Fig. 4	**10a**	Fig. S18	S55
Fig. 4	**10b**	Fig. S19	S56
Fig. 4	**10c**	Fig. S20	S57
Fig. 4	**10d**	Fig. S6	S12
Fig. 4	**10f**	Fig. S11	S44, S45
Fig. 5	**10f**	Fig. S11, Fig. S5a	S43, S44, S11
Fig. 5	**10g**	Fig. S12, Fig. S5b	S46, S11
Fig. 5	**10h**	Fig. S13, Fig. S5c	S47, S48, S11
Fig. 5	**10i**	Fig. S14, Fig. S5d	S49, S50, S11
Fig. 5	**10j**	Fig. S15, Fig. S5e, Sec. 10	S51, S52, S11
Fig. 5	**10k**	Fig. S16, Fig. S5f	S52, S53, S11
Fig. 5	**10l**	Fig. S17, Fig. S5g	S54, S11
Fig. 6	**10f**	Fig. S11	S44, S44
Fig. 6	**10m**	Fig. S28	S65
Fig. 6	**13**	Fig. S10, Fig. S28, Fig. S29	S14, S65
Fig. S2	**7a**	Sec. 6	S20
Fig. S8	**S16, S17**	Fig. S30	S66
Fig. S9	**S18**, **S19**, **S20**	Fig. S31, Fig. S32	S67
Table S1	**7a**	Sec. 6	S20
Table S5	**5a**	Sec. 6	S28
Table S5	Mono-labeled RNase A	Sec. 7.2 – Fig. S18	S55
Table S5	Mono-, bis-labeled RNase A	Sec. 7.2 – Fig. S19	S56
Table S5	Mono-, bis-, tris-labeled RNase A	Sec. 7.2 – Fig. S20	S57
Table S5	Mono-labeled Melittin	Sec. 7.2 – Fig. S21	S58
Table S5	Mono-, bis-, tris-labeled Melittin	Sec. 7.2 – Fig. S22	S59
Table S5	Mono-labeled Ubiquitin	Sec. 7.2 – Fig. S23	S60
Table S5	Mono-, bis-, tris-labeled Ubiquitin	Sec. 7.2 – Fig. S24	S61
Table S5	Mono-labeled Myoglobin	Sec. 7.2 – Fig. S25	S62
Table S5	Mono-, bis-labeled Myoglobin	Sec. 7.2 – Fig. S26	S63
Table S5	Mono-labeled Histone (H 3.1)	Sec. 7.2 – Fig. S27	S64
Fig. S37	**9f**, **10f**	Sec. 9.1, Fig. S37	S69, S70
Fig. S38	**9i**, **10i**	Sec. 9.1, Fig. S38	S70, S71
Fig. S39	**9f**, **9h**, **9i**, **9j**, **9k**, **10f**, **10h**, **10i**, **10j**	Sec. 10, Fig. S39	S71

5. Additional results and discussion

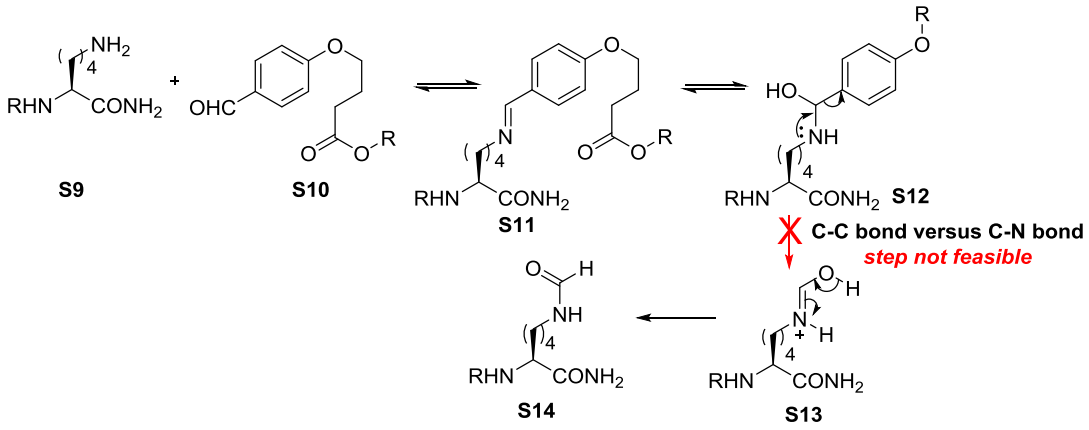


Figure S1. A hypothetical pathway for direct formylation of primary amine **S9** through an aldehyde **S10**.

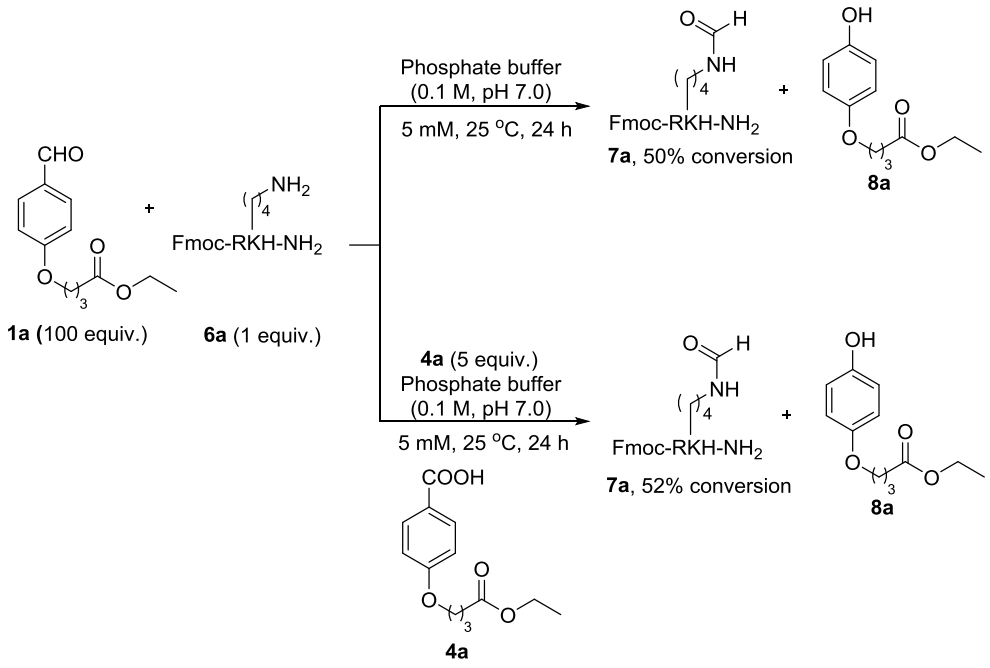


Figure S2. The role of carboxylic acid on formylation of the peptide. Carboxylic acid as additive displayed no effect on the transformation.

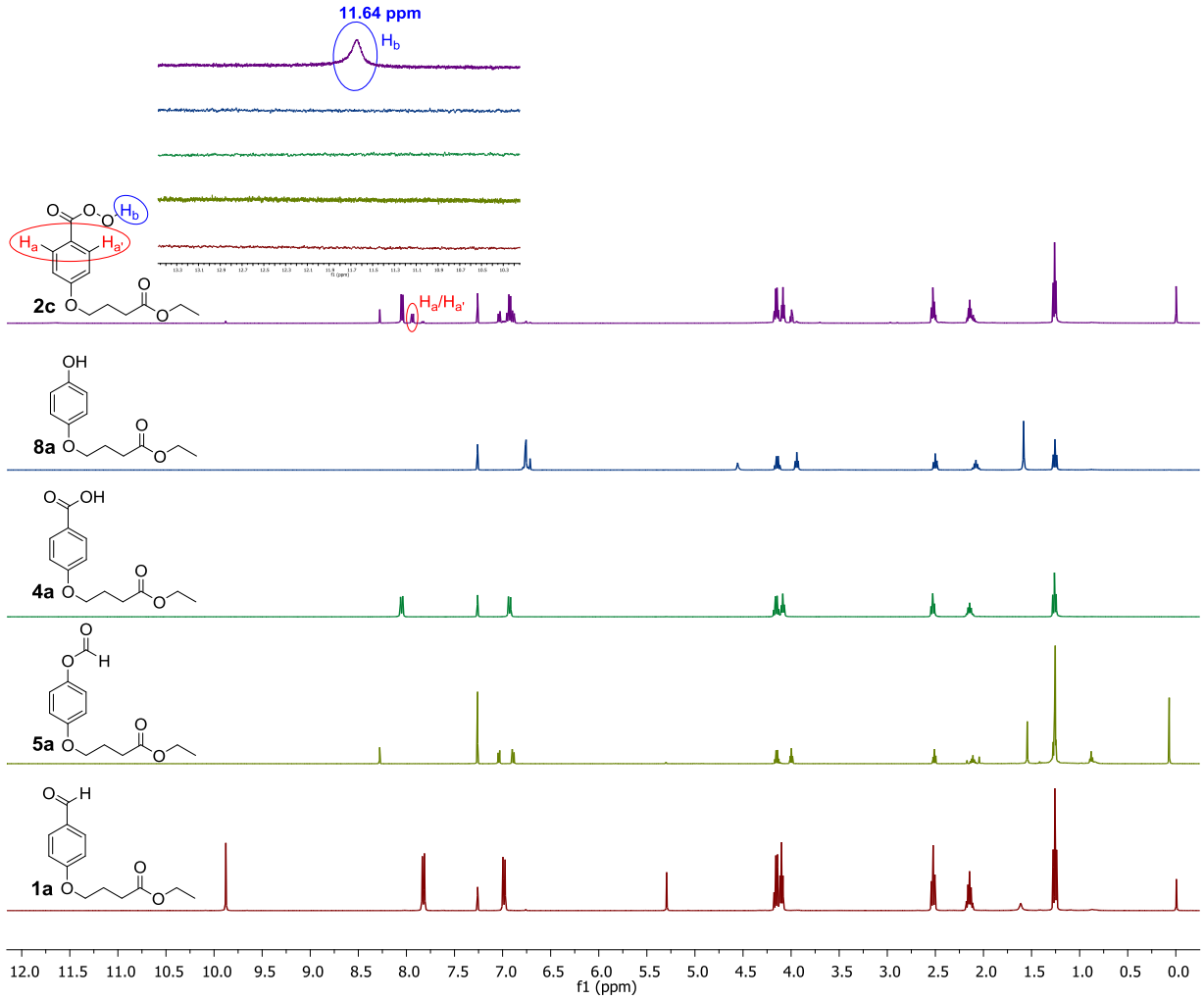
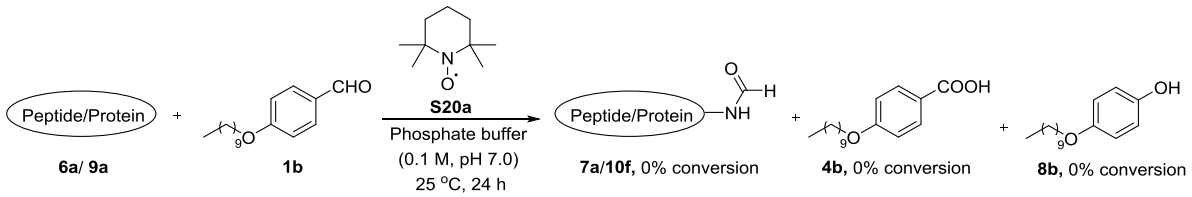


Figure S3. Comparative analysis of crude reaction mixture with intermediates in proposed reaction pathway by ^{1}H NMR spectroscopy. Identification of the key intermediate, i.e. peracid **2c**. Deconvoluted aromatic protons ($H_a/H_{a'}$, 7.94 ppm, 2H) and the proton from peracid (H_b, 11.64 ppm, 1H) of **2c** are highlighted in red and blue circles respectively.



Case a: Peptide, Fmoc-RKH-NH$_2$ (**6a**, 5 mM), aldehyde (**1b**, 1 M), TEMPO (**S20a**, 0.6 M).

Case b: Protein, RNase A (**9a**, 73 μM), aldehyde (**1b**, 43.8 mM), TEMPO (**S20a**, 8.76 mM).

Figure S4. **Effect of radical quencher on formylation**. The complete inhibition of the reaction supports the critical role of radical species in the generation of the key intermediate (peracid, **2c**). % Conversions were monitored using LC-MS.

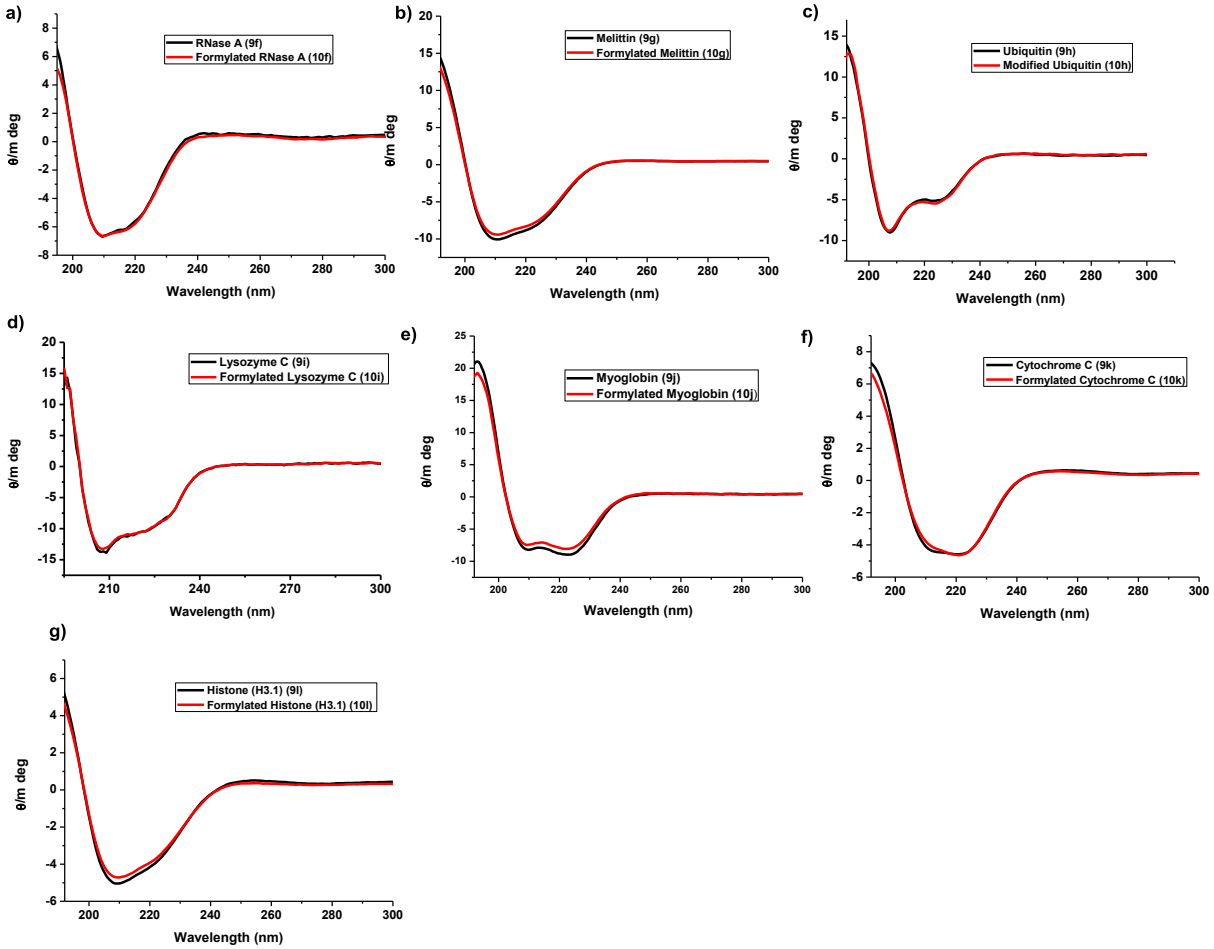


Figure S5. Effect of formylation on the structure of proteins. a) Circular Dichroism (CD) spectra of RNase A **9f** (black line), formylated RNase A **10f** (red line) in phosphate buffer (0.1 M, pH 7.0) at concentration 0.1 mg/ml. b) Melittin **9j** (black line), formylated Melittin **10j** (red line) in phosphate buffer (0.1 M, pH 7.0) at concentration 0.063 mg/ml. c) Ubiquitin **9h** (black line), formylated Ubiquitin **10h** (red line) in phosphate buffer (0.1 M, pH 7.0) at concentration 0.186 mg/ml. d) Circular Dichroism (CD) spectra of Lysozyme C **9i** (black line), formylated Lysozyme C **10i** (red line) in phosphate buffer (0.1 M, pH 7.0) at concentration 0.1mg/ml. e) Circular Dichroism (CD) spectra of Myoglobin **9j** (black line), formylated Myoglobin **10j** (red line) in phosphate buffer (0.1 M, pH 7.0) at concentration 0.12 mg/ml. f) Circular Dichroism (CD) spectra of Cytochrome C **9k** (black line), formylated Myoglobin **10k** (red line) in phosphate buffer (0.1 M, pH 7.0) at concentration 0.09 mg/ml. g) Circular Dichroism (CD) spectra of Histone (H 3.1) **9i** (black line), formylated Histone (H 3.1) **10i** (red line) in phosphate buffer (0.1 M, pH 7.0) at concentration 0.05 mg/ml. The concentration of formylated protein is kept constant with respect to the native protein in each case.

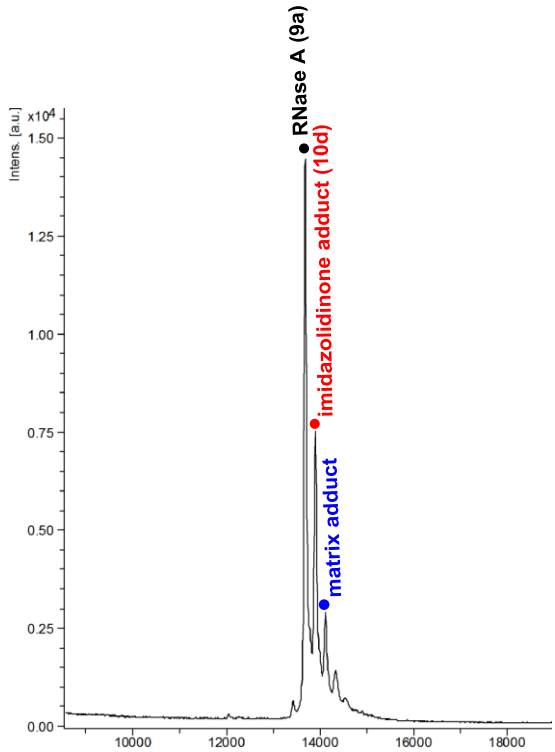


Figure S6. MS spectra for imidazolidinone **10d** with RNase A **9a** and aldehyde **1a**.

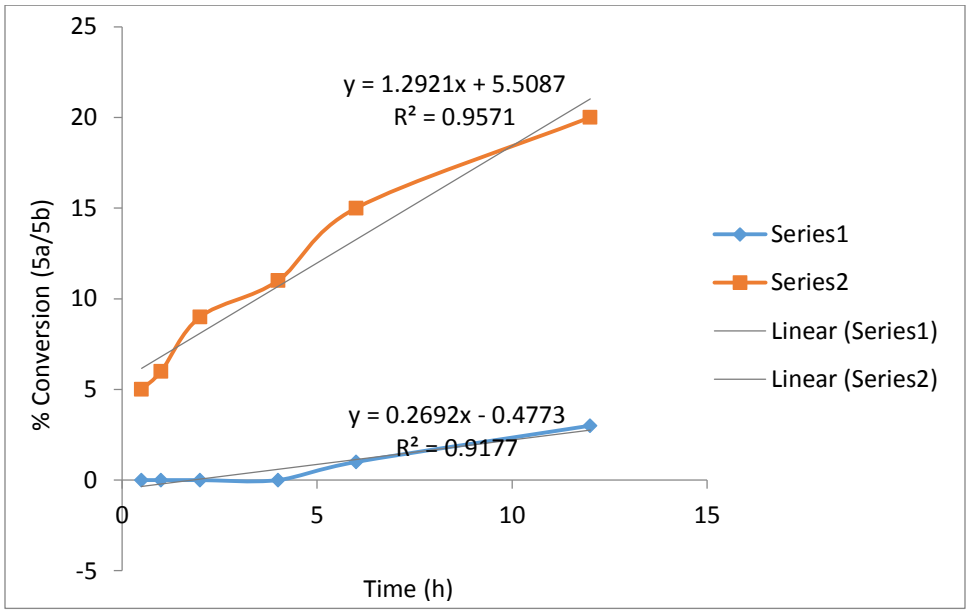


Figure S7. Effect of hydrophobic residue on formate ester (**5a**/**5b**) generation in phosphate buffer (0.1 M, pH 7.0). Series 1: **5a** (**1a**, 73 µM), Series 2: **5b** (**1b**, 73 µM). At higher concentration of **1a**/**1b** (73 × 600 µM), 1% conversion is observed for **5a** and 15% conversion for **5b** over a period of 24 h (not shown in the plot). The aldehyde **1a** results in ~5-6 equivalents of formate **5a** over a period of 48 hours under the given reaction conditions for protein formylation.

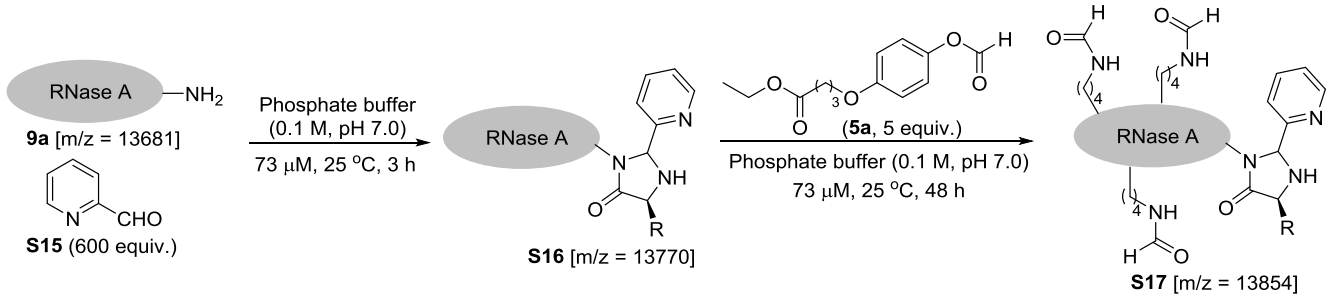


Figure S8. Formylation of RNase A (**9a**) with pre-synthesized formate ester **5a** in the presence of pyridine 2-carboxaldehyde (*refer Figure S30*).

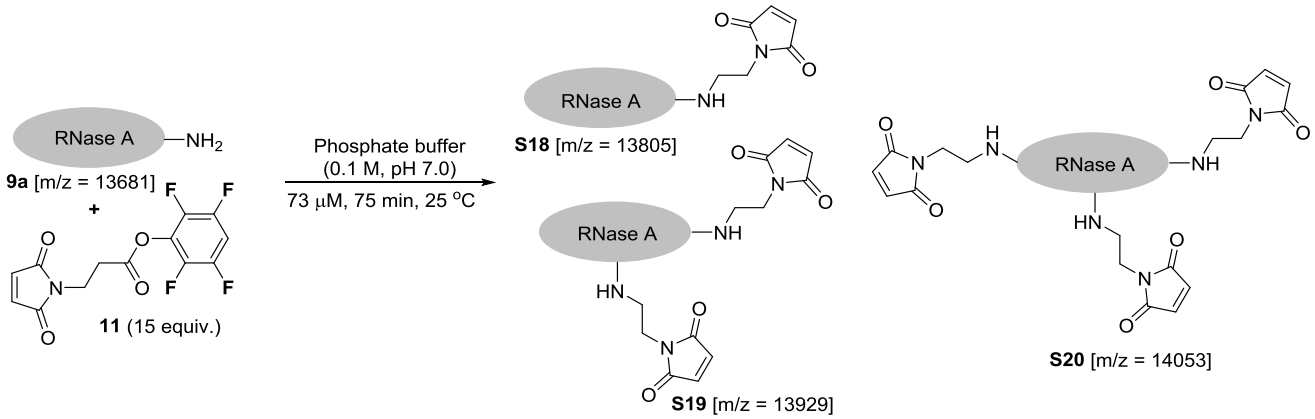


Figure S9. Acylation of RNase A **9a** in the absence of aldehyde **1a** (*refer Figure S31 and S32*).

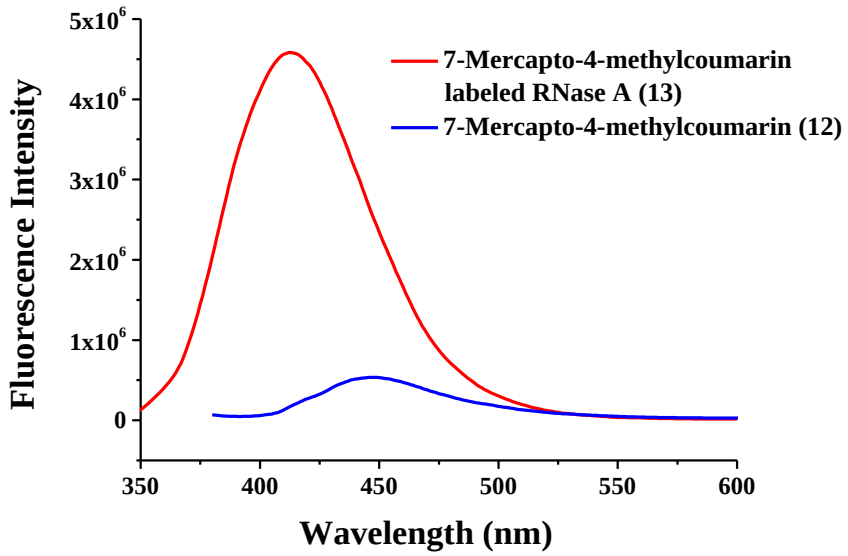
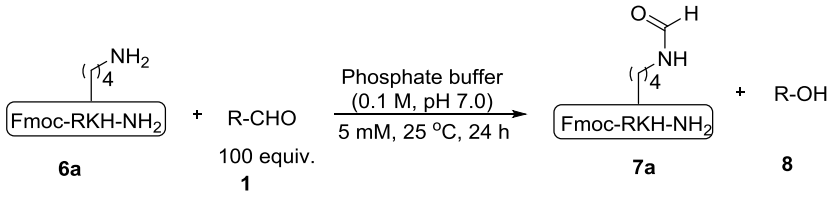
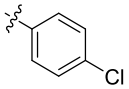
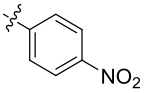
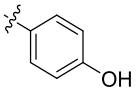
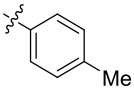
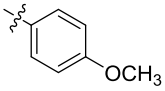
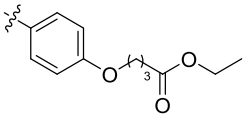
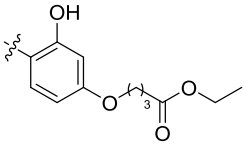
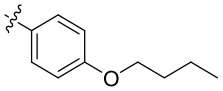
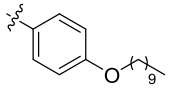


Figure S10. Steady-state fluorescence spectra of 7-mercapto-4-methylcoumarin **12** (blue line) and 7-mercapto-4-methylcoumarin labeled RNase A **13** (red line) in phosphate buffer (0.1 M, pH 7.0). In aqueous buffer, **12** exhibit absorption and emission band peaked at 365 nm and 447 nm, respectively. In the emission profile of **13**, the fluorescence intensity increased drastically and centered at 412 nm, thereby showing a blue shift of 35 nm. The emission data is a clear signature of the increased hydrophobic environment experienced by the fluorophore when covalently attached to the protein as compared to the bulk. [*For characterization of* **13**, *also see MS data and peptide mapping in Supplementary Figures S28 and S29*]

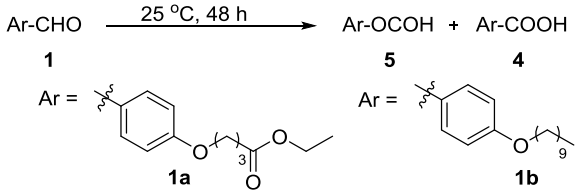
Table S1. The role of aldehydes **1** as formylating pre-reagent.



Entry[a]	R	% Conversion 7a[b]
1		<5
2		0
3		0
4		<5
5		0
6		0
7		0
8		18
9		33
10		50
11		0
12		45
13		89

[a] Peptide (1 equiv.) and aldehyde (100 equiv.) used in reactions. [b] % Conversions were monitored by using LC-MS.

Table S2. Parameters regulating generation of formate **5**.

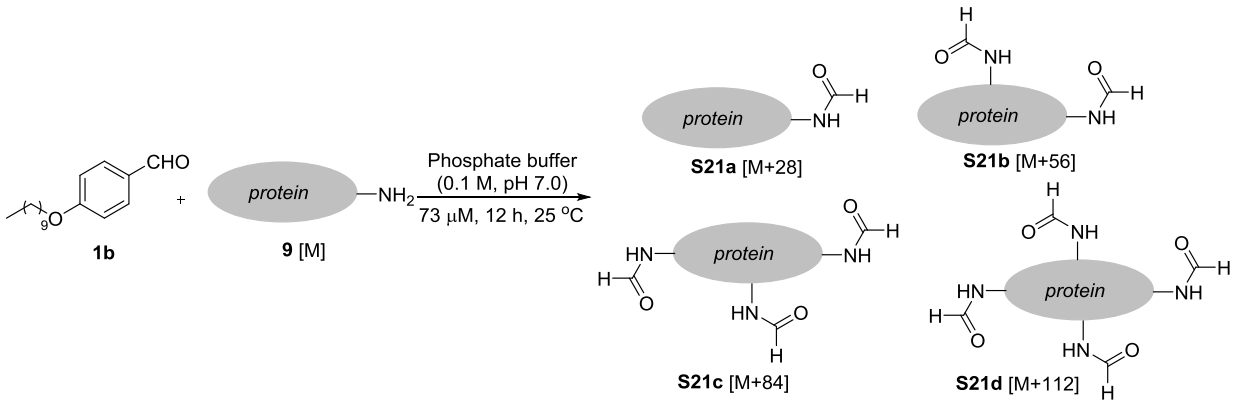


Entry	Aldehyde	Conditions	% Conversion[c] (5a/5b)	% Conversion[c] (4a/4b)
1[a]	**1b**	Ambient	27 (**5b**)	24 (**4b**)
2[a]	**1b**	O_2	30 (**5b**)	25 (**4b**)
3[a]	**1b**	N_2	0 (**5b**)	0 (**4b**)
4[b]	**1b**	N_2	0 (**5b**)	0 (**4b**)
5[b]	**1b**	Ambient	5 (**5b**)	5 (**4b**)
6[b]	**1a**	N_2	0 (**5a**)	0 (**4a**)
7[b]	**1a**	Ambient	4 (**5a**)	20 (**4a**)

[a] Aldehyde **1b** stored in ambient solvent free conditions. [b] **1b** (5 × 200 mM), Phosphate buffer (pH 7.0, 0.1 M). [c] % Conversion analyzed based on ^{1}H-NMR of the crude reaction mixture.

In solvent free ambient conditions, the pure aldehyde **1b** led to the formation of 27-30% formate **5a** in 48 h in ambient conditions or under the oxygen atmosphere (Table S2). Upon incubation under nitrogen or argon, the aldehyde was chemically intact over a period of ten days. If the reaction mixture containing Fmoc-RKH-NH_2 **6a** and aldehyde **1b** are deprived of oxygen or light or both, it results in 25-31% conversions (98% conversion in ambient conditions). The low conversions can be due to the presence of traces of oxygen that can result in formate **5b** below detection levels of ^{1}H NMR spectroscopy.

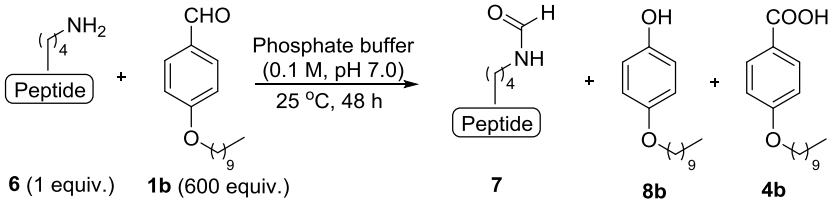
Table S3. Optimization of stoichiometry.



Entry	**Protein** [MW in Da]	**Conversion**[a] **1b (300 equiv.)**	**Conversion**[a] **1b (600 equiv.)**
1	RNase A[b, c] [13681]	**S21a** (40%)	**S21a** (82%)
2	Ubiquitin[c] [8561]	**S21a** (35%)	**S21a**, **S21b**, **S21c**
3	Lysozyme C[c] [14304]	**S21a** (35%)	**S21a**, **S21b**, **S21c**
4	Myoglobin[c] [16951]	**S21a**, **S21b**	**S21a**, **S21b**, **S21c**
5	Cytochrome C[c] [12360]	**S21a** (45%), **S21b** (10%)	**S21a**, **S21b**, **S21c**, **S21d**

[a] % Conversions were monitored using LC-MS and MALDI-ToF-MS. Formylation was not accompanied by any side reaction. [b] **1b** (1000 equiv.) resulted in mixture of **S21a** and **S21b**. [c] The optimum stoichiometry of formylation pre-reagent **1b** was found to be 600 equivalents for RNase A. **S21a**, **S21b**, **S21c** and **S21d** are mono-, bis-, tris- and tetra-labeled proteins respectively. For all the other proteins (entries 2-5), 600 equivalents of **1a** resulted in high conversions retaining the selectivity. In the case of Melittin, 300 equivalents of **1a** was appropriate for achieving high reactivity and selectivity.

Table S4. Analysis of side reactions with aldehyde **1b** at varying concentration of peptides.

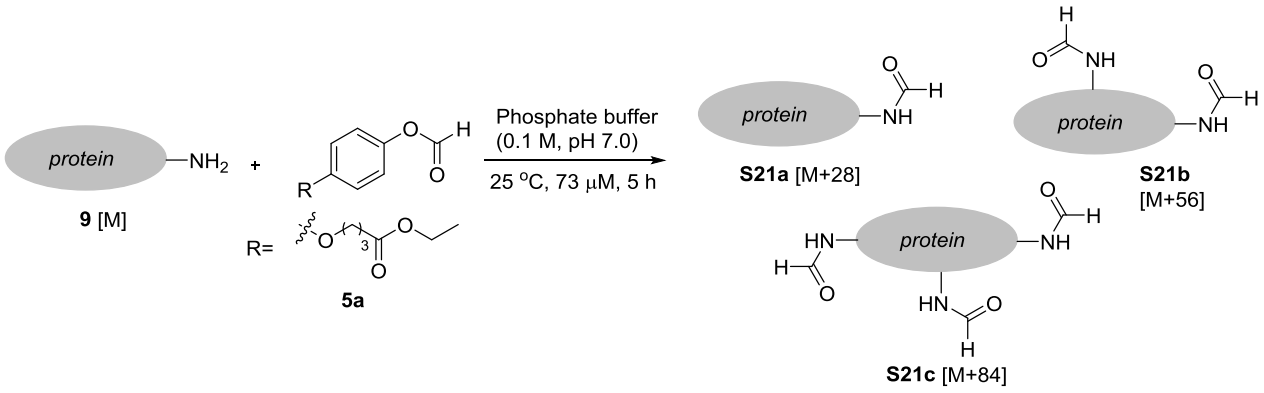


6 (1 equiv.) **1b** (600 equiv.) **7** **8b** **4b**

Entry	Peptide	Concentration	%Conversion (7)[a]
1	Fmoc-RKH-NH_2	100 μM	85
		1000 μM	>99
2	Fmoc-WK-NH_2	73 μM	40
		100 μM	50
		1000 μM	>99
3	Fmoc-YK-NH_2	1000 μM	>99
4	Fmoc-TK-NH_2	73 μM	>99
		100 μM	>99
		1000 μM	>99

[a] %Conversion were analyzed by LC-MS. No side reactions were observed in entries 1-4.

Table S5. Kinetic reactivity order of Lys residues in formylation of proteins using pre-synthesized formate ester **5a**.



Entry	**Protein** [MW in Da]	**Formate ester (5a, equiv.)**	**Conversion**[a]	**Site(s) of modification**[b]
1	**RNase A** [13681]	5	**S21a** (35%), **S21b**, **S21c**	K1, K37, K91
2	**Melittin** [2846]	5	**S21a** (45%), **S21b** (36%), **S21c**	G1, K21, K23
3	**Ubiquitin** [8561]	3	**S21a** (40%)	M1
		5	**S21a** (45%), **S21b** (15%), **S21c**	M1, K48, K63
4	**Lysozyme C** [14304]	5	**S21a** (40%)	K116
5	**Myoglobin** [16951]	3	**S21a** (30%)	K145
		5	**S21a** (40%), **S21b**	K145, K133
6	**Histone (H 3.1)** [17442]	5	**S21a** (50%)	K37

[a] %Conversions were analyzed by LC-MS and MALDI-ToF-MS. [b] Site of modification was analyzed by MS and MS-MS. **S21a**, **S21b** and **S21c** are mono-, bis- and tris-labeled proteins respectively [*refer Section 7.2 (figure S18- S27)*].

6. Characterization data

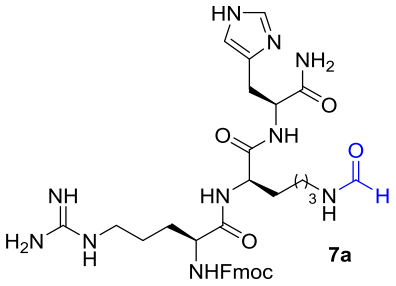
NMR: ^{1}H, ^{13}C, COSY and HSQC

HPLC (peak identification by MS): Acetonitrile and H_2O were buffered with 0.01% formic acid.

Method A (Column: ZORBAX 300SB-C18 5 µM 4.6 × 250 mm)		
Time (min)	**Acetonitrile (%)**	**H_2O (%)**
0	10	90
7	10	90
35	70	30
38	90	10
40	10	90
45	10	90

Method B (Column: Poroshell 120 EC-C18 2.7 µM 3.0 × 50 mm)		
Time (min)	**Acetonitrile (%)**	**H_2O (%)**
0	10	90
2	50	50
8	60	40
12	90	10

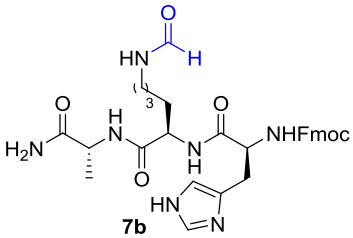
(9H-fluoren-9-yl)methyl((S)-1-(((R)-1-(((S)-1-amino-3-(1H-imidazol-4-yl)-1-oxopropan-2-yl)amino)-6-formamido-1-oxohexan-2-yl)amino)-5-guanidino-1-oxopentan-2-yl)carbamate (7a)



7a

Yield 89%, TLC (methanol:chloroform 3:7, R_f 0.30), HPLC (Method A, t_R = 24.74 min), ^{1}H NMR (700 MHz, DMSO-d_6) δ 8.12 (m, 2H), 8.05 (m, 1H), 7.98 (s, 1H), 7.90 (d, *J* = 7.5 Hz, 1H), 7.72 (dd, *J* = 16.0, 7.5 Hz, 1H), 7.61 (d, *J* = 7.7 Hz, 1H), 7.53 (s, 1H), 7.46-7.40 (m, 1H), 7.34-7.31 (m, 1H), 4.56 (dd, *J* = 14.4, 7.2 Hz, 1H), 4.38 (d, *J* = 6.1 Hz, 1H), 4.28 (dd, *J* = 14.5, 8.9 Hz, 1H), 4.25-4.20 (m, 1H), 4.18 (s, 1H), 4.05 (s, 1H), 3.13-2.99 (m, 3H), 2.93 (d, *J* = 37.8 Hz, 2H), 2.82 (dd, *J* = 14.5, 8.3 Hz, 1H), 1.66 (d, *J* = 37.3 Hz, 2H), 1.52 (s, 3H), 1.42-1.31 (m, 2H), 1.21 (d, *J* = 37.1 Hz, 3H). ^{13}C NMR (175 MHz, DMSO-d_6) δ 173.0, 172.5, 172.0, 161.4, 157.1, 156.4, 144.3, 144.2, 141.1, 128.1, 127.5, 125.7, 120.6, 66.1, 54.2, 53.8, 53.1, 50.6, 47.1, 37.2, 31.3, 28.96, 25.2, 24.9, 23.0, 21.5 ppm. HRMS (ESI) $[MH]^+$ calcd. $C_{34}H_{44}N_{10}O_6$ 689.3524, found 689.3518.

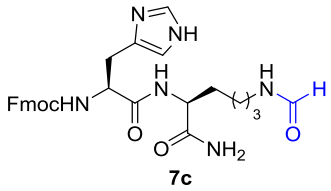
(9H-fluoren-9-yl)methyl((S)-1-(((S)-1-(((S)-1-amino-1-oxopropan-2-yl)amino)-6-formamido-1-oxohexan-2-yl)amino)-3-(1H-imidazol-4-yl)-1-oxopropan-2-yl)carbamate (7b)



7b

Yield 79%, TLC (methanol:chloroform 1:9, R_f 0.20), HPLC (Method A, t_R = 24.74 min), ^{1}H NMR (700 MHz, MeOH-d_4) δ 7.90 (s, 1H), 7.79 (s, 1H), 7.70 (d, *J* = 7.5 Hz, 2H), 7.52 (t, *J* = 7.3 Hz, 2H), 7.29 (t, *J* = 7.3 Hz, 2H), 7.21 (t, *J* = 6.2 Hz, 2H), 6.88 (s, 1H), 4.50 (s, 2H), 4.33-4.26 (m, 2H), 4.22 (dd, *J* = 13.2, 8.6 Hz, 2H), 4.20-4.16 (m, 1H), 4.10 (t, *J* = 6.6 Hz, 1H), 3.10 (s, 2H), 3.01 (dd, *J* = 15.0, 5.3 Hz, 1H), 2.89 (dd, *J* = 14.9, 7.3 Hz, 1H), 1.73 (s, 2H), 1.59 (d, *J* = 8.8 Hz, 2H), 1.46-1.36 (m, 2H) ppm. ^{13}C NMR (175 MHz, MeOH-d_4) δ 177.6, 174.0, 174.1, 163.9, 158.4, 145.3, 145.2, 142.7, 136.2, 128.9, 128.3, 126.3, 126.2, 121.0, 68.1, 56.3, 55.1, 50.2, 48.4, 38.6, 31.1, 29.9, 27.0, 23.9, 18.2 ppm. HRMS (ESI) $[MH]^+$ calcd. $C_{31}H_{37}N_7O_6$ 604.2884, found 604.2961.

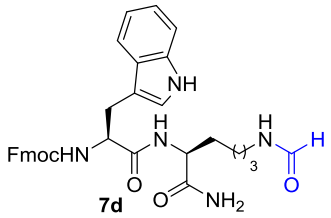
(9H-fluoren-9-yl) methyl ((R)-1-(((S)-1-amino-6-formamido-1-oxohexan-2-yl) amino)-3-(1H-imidazol-5-yl)-1-oxopropan-2-yl) carbamate (7c)



7c

Yield 94%, HPLC (Method B, t_R = 3.05 min), ^{1}H NMR (700 MHz, DMSO-d_6) δ 8.06 (s, 1H), 7.99 (s, 1H), 7.97 (s, 1H), 7.88 (d, *J* = 6.3 Hz, 2H), 7.68 (d, *J* = 7.5 Hz, 2H), 7.41 (t, *J* = 7.2 Hz, 2H), 7.33 (dd, *J* = 12.2, 7.0 Hz, 2H), 4.24 (s, 1H), 4.21 (t, *J* = 9.1 Hz, 3H), 4.15-4.09 (m, 1H), 3.16 (d, *J* = 4.3 Hz, 2H), 3.06-2.98 (m, 2H), 1.72-1.62 (m, 1H), 1.52 (d, *J* = 9.5 Hz, 1H), 1.43-1.31 (m, 2H), 1.19 (ddd, *J* = 22.5, 13.0, 8.1 Hz, 2H) ppm. ^{13}C NMR (175 MHz, DMSO-d_6) δ 173.7, 171.1, 161.1, 155.7, 143.8, 143.7, 140.7, 134.7, 127.7, 127.2, 125.4, 125.3, 120.1, 65.8, 65.1, 54.9, 52.4, 46.6, 40.1, 36.9, 28.6, 22.5 ppm. MS (ESI) $[MH]^+$ calcd. $C_{28}H_{32}N_6O_5$ 533.2, found 533.3.

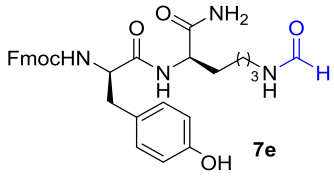
(9H-fluoren-9-yl) methyl ((S)-1-(((S)-1, 6-diamino-1-oxohexan-2-yl) amino)-3-(1H-indol-3-yl)-1-oxopropan-2-yl) carbamate (7d)



7d

Yield 80%, TLC (methanol:chloroform 1:9, R_f 0.35), HPLC (Method B, t_R = 4.66 min), ^{1}H NMR (700 MHz, DMSO-d_6) δ 7.90-7.85 (m, 2H), 7.66 (d, *J* = 7.8 Hz, 1H), 7.64 (d, *J* = 7.5 Hz, 1H), 7.61 (d, *J* = 7.5 Hz, 1H), 7.57 (d, *J* = 8.3 Hz, 1H), 7.39 (dt, *J* = 19.1, 8.0 Hz, 2H), 7.31 (dd, *J* = 17.5, 7.6 Hz, 2H), 7.25 (t, *J* = 7.4 Hz, 1H), 7.06 (dd, *J* = 15.3, 7.8 Hz, 2H), 4.31 (td, *J* = 9.6, 4.3 Hz, 1H), 4.20 (dt, *J* = 16.6, 8.3 Hz, 1H), 4.15 (s, 2H), 4.06 (dt, *J* = 14.1, 7.0 Hz, 1H), 3.13 (dd, *J* = 14.5, 3.7 Hz, 1H), 3.02 (dt, *J* = 18.8, 6.5 Hz, 2H), 2.94 (dd, *J* = 14.5, 10.2 Hz, 1H), 1.65 (dt, *J* = 27.8, 16.0 Hz, 2H), 1.58-1.48 (m, 2H), 1.37 (dd, *J* = 13.3, 7.1 Hz, 2H) ppm. ^{13}C NMR (175 MHz, DMSO-d_6) δ 173.9, 172.1, 161.3, 156.2, 144.2, 144.1, 141.1, 136.5, 128.0, 127.7, 127.5, 125.8, 125.7, 124.2, 121.3, 120.5, 111.7, 110.7, 66.1, 56.0, 52.7, 47.01, 37.4, 32.0, 29.2, 23.0, 22.5 ppm. MS (ESI) $[MH]^+$ calcd. $C_{33}H_{35}N_5O_5$ 582.2, found 582.3.

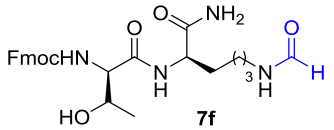
(9H-fluoren-9-yl)methyl ((S)-1-(((S)-1-amino-6-formamido-1-oxohexan-2-yl)amino)-3-(4-hydroxyphenyl)-1-oxopropan-2-yl)carbamate (7e)



7e

Yield 80%, TLC (methanol:chloroform 1:9, R_f 0.25), HPLC (Method B, t_R = 4.35 min), ^{1}H NMR (700 MHz, DMSO-d_6) δ 7.83 (d, *J* = 5.4 Hz, 2H), 7.62 – 7.55 (m, 2H), 7.36 (s, 2H), 7.26 (s, 2H), 7.03 (d, *J* = 6.3 Hz, 2H), 6.58 (d, *J* = 6.5 Hz, 2H), 4.13 (s, 2H), 4.11 (s, 1H), 4.09 (s, 2H), 2.98 (s, 2H), 2.85 (d, *J* = 13.6 Hz, 1H), 2.60 (t, *J* = 11.8 Hz, 1H), 1.59 (s, 1H), 1.47 (s, 1H), 1.30 (d, *J* = 48.2 Hz, 2H), 1.20 (d, *J* = 17.3 Hz, 2H) ppm. ^{13}C NMR (175 MHz, DMSO-d_6) δ 173.4, 171.4, 160.9, 155.7, 143.8, 143.7, 130.1, 128.2, 127.6, 127.1, 125.3, 120.1, 114.8, 65.6, 56.5, 52.2, 46.5, 37.0, 36.6, 31.8, 28.7, 22.6. HRMS (ESI) $[MH]^+$ calcd. $C_{31}H_{34}N_4O_6$ 559.2557, found 559.2551.

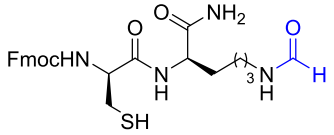
(9H-fluoren-9-yl) methyl ((2S, 3R)-1-(((S)-1-amino-6-formamido-1-oxohexan-2-yl) amino)-3-hydroxy-1-oxobutan-2-yl) carbamate (7f)



7f

Yield 86%, HPLC (Method B, t_R = 6.31 min), ^{1}H NMR (700 MHz, DMSO-d_6) δ 7.89 (d, *J* = 7.6 Hz, 2H), 7.75 (dd, *J* = 11.8, 7.6 Hz, 2H), 7.45-7.39 (m, 2H), 7.33 (t, *J* = 7.3 Hz, 2H), 4.31 (dd, *J* = 10.2, 7.4 Hz, 1H), 4.26 (dd, *J* = 16.6, 6.3 Hz, 1H), 4.23 (t, *J* = 7.0 Hz, 1H), 4.18 (dd, *J* = 13.2, 8.3 Hz, 1H), 4.01 (dd, *J* = 8.6, 4.4 Hz, 1H), 3.99-3.94 (m, 1H), 3.03 (td, *J* = 12.5, 6.8 Hz, 2H), 1.67 (ddd, *J* = 14.5, 10.8, 5.6 Hz, 1H), 1.56-1.48 (m, 1H), 1.43-1.34 (m, 2H), 1.28 (dd, *J* = 23.0, 8.0 Hz, 2H), 1.05 (d, *J* = 7.3 Hz, 3H) ppm. ^{13}C NMR (175 MHz, DMSO-d_6) δ 174.0, 170.2, 161.38, 156.6, 144.3, 144.2, 141.1, 128.1, 127.5, 125.8, 120.5, 67.3, 66.2, 52.7, 47.1, 37.43, 32.0, 29.1, 23.0, 20.0 ppm. MS (ESI) $[MH]^+$ calcd. $C_{26}H_{32}N_4O_6$ 497.2, found 497.2.

(9H-fluoren-9-yl)methyl(S)-1-(R)-1-amino-6-formamido-1-oxohexan-2-yl)amino)-3-mercapto-1-oxopropan-2-yl)carbamate (7g)

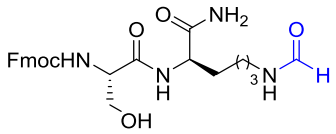


7g

Yield 82%, HPLC (Method A, t_R = 8.23 min), ^{1}H NMR (700 MHz, DMSO-d_6) δ 7.90 (d, *J* = 7.6 Hz, 2H), 7.76-7.70 (m, 2H), 7.42 (t, *J* = 7.4 Hz, 2H), 7.34 (d, *J* = 7.5 Hz, 2H), 4.29 (ddd, *J* = 23.7, 10.4, 7.3 Hz, 2H), 4.23 (dd, *J* = 18.5, 11.8 Hz, 1H), 4.14 (dt, *J* = 13.2, 6.6 Hz, 2H), 3.06-2.99 (m, 2H), 2.82 (ddd, *J* = 13.4, 8.6, 4.8 Hz, 1H), 2.67 (dt, *J* = 13.5, 8.4 Hz, 1H), 1.69-1.60 (m, 1H), 1.51 (dt, *J* = 18.5, 11.6 Hz, 1H),

1.42-1.32 (m, 2H), 1.31-1.20 (m, 2H) ppm. ^{13}C NMR (175 MHz, DMSO-d_6) δ 173.8, 170.3, 161.3, 156.4, 144.3, 144.2, 128.1, 127.5, 125.7, 120.6, 66.2, 57.6, 52.8, 47.0, 37.4, 32.0, 29.1, 26.7, 23.0 ppm. MS (ESI) $[MH]^+$ calcd. $C_{25}H_{30}N_4O_5S$ 499.1, found 499.2.

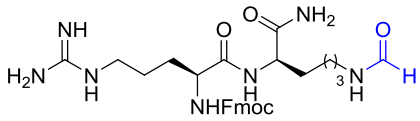
(9H-fluoren-9-yl)methyl(S)-1-(S)-1-amino-6-formamido-1-oxohexan-2-yl)amino)-3-hydroxy-1-oxopropan-2-yl)carbamate (7h)



7h

Yield 81%, HPLC (Method B, t_R = 3.42 min), ^{1}H NMR (700 MHz, DMSO-d_6) δ 7.84 (d, J = 7.5 Hz, 2H), 7.68 (dd, J = 9.9, 7.9 Hz, 2H), 7.37 (t, J = 7.6 Hz, 2H), 7.28 (t, J = 7.6 Hz, 2H), 4.25-4.20 (m, 2H), 4.17 (t, J = 7.0 Hz, 1H), 4.09 (td, J = 8.6, 4.8 Hz, 1H), 4.05 (dd, J = 14.1, 6.2 Hz, 1H), 3.57-3.49 (m, 2H), 2.97 (td, J = 13.1, 6.7 Hz, 2H), 1.64 (ddd, J = 14.0, 10.6, 5.7 Hz, 1H), 1.51-1.41 (m, 1H), 1.38-1.28 (m, 2H), 1.27-1.14 (m, 2H) ppm. ^{13}C NMR (175 MHz, DMSO-d_6) δ 174.0, 170.6, 161.3, 156.3, 144.2, 141.1, 128.1, 127.5, 125.7, 120.6, 66.2, 62.3, 57.2, 52.8, 47.0, 37.4, 31.7, 29.1, 23.1 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_{25}H_{30}N_4O_6$ 505.2063, found 505.2058.

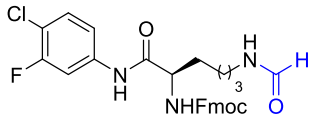
(9H-fluoren-9-yl)methyl ((R)-1-(((S)-1-amino-6-formamido-1-oxohexan-2-yl)amino)-5-guanidino-1-oxopentan-2-yl)carbamate (7i)



7i

Yield 78%, TLC (methanol:chloroform 1:9, R_f 0.35), HPLC (Method B, t_R = 4.60 min), ^{1}H NMR (700 MHz, DMSO-d_6) δ 8.11 (d, J = 9.1 Hz, 1H), 7.88 (t, J = 16.1 Hz, 2H), 7.73 (dd, J = 15.2, 7.4 Hz, 2H), 7.41 (t, J = 7.4 Hz, 2H), 6.92 (d, J = 9.1 Hz, 1H), 4.27 (d, J = 7.0 Hz, 1H), 4.24 -4.19 (m, 1H), 4.14 (dd, J = 13.6, 8.3 Hz, 1H), 4.02 (dd, J = 14.1, 7.0 Hz, 1H), 3.14-2.97 (m, 4H), 1.72 (s, 1H), 1.67-1.43 (m, 6H), 1.37 (t, J = 29.4 Hz, 2H), 1.31-1.24 (m, 2H) ppm. ^{13}C NMR (175 MHz, DMSO-d_6) δ 173.7, 171.5, 161.0, 156.7, 156.0, 143.9, 143.7, 127.6, 127.1, 125.3, 120.1, 64.9, 54.2, 52.5, 46.6, 40.2, 36.9, 31.5, 28.6, 25.0, 22.6, 20.8 ppm. HRMS (ESI) $[MH]^+$ calcd. $C_{28}H_{37}N_7O_5$ 552.2934, found 552.2929.

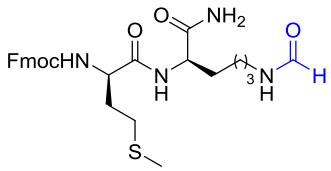
(9H-fluoren-9-yl) methyl (S)-(1-((3-chloro-4-fluorophenyl) amino)-6-formamido-1-oxohexan-2-yl) carbamate (7j)



7j

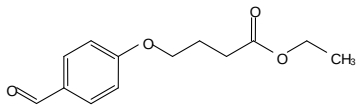
Yield 93%, HPLC (Method B, t_R = 4.76 min), ^{1}H NMR (700 MHz, DMSO-d_6) δ 7.93 (dd, J = 6.7, 2.2 Hz, 1H), 7.89 (d, J = 7.5 Hz, 2H), 7.76-7.68 (m, 3H), 7.41 (t, J = 6.8 Hz, 2H), 7.37 (dd, J = 16.9, 7.9 Hz, 1H), 7.33 (dd, J = 13.3, 6.9 Hz, 2H), 4.31-4.25 (m, 2H), 4.22 (t, J = 7.1 Hz, 1H), 4.07 (td, J = 8.8, 5.4 Hz, 1H), 3.10-3.03 (m, 2H), 1.72-1.57 (m, 2H), 1.47-1.34 (m, 3H), 1.34-1.25 (m, 1H) ppm. ^{13}C NMR (175 MHz, DMSO-d_6) δ 171.4, 160.9, 156.1, 153.7, 152.4, 143.8, 143.8, 140.7, 136.2, 127.6, 127.1, 125.3, 120.5, 120.1, 116.9, 65.6, 55.4, 46.6, 36.9, 31.2, 28.7, 23.1 ppm. HRMS (ESI) $[M+Na]^+$ calcd. $C_{28}H_{27}ClFN_3O_4$ 546.1572, found 546.1566.

(9H-fluoren-9-yl)methyl ((S)-1-(((S)-1-amino-6-formamido-1-oxohexan-2-yl)amino)-4-(methylthio)-1-oxobutan-2-yl)carbamate (7k)

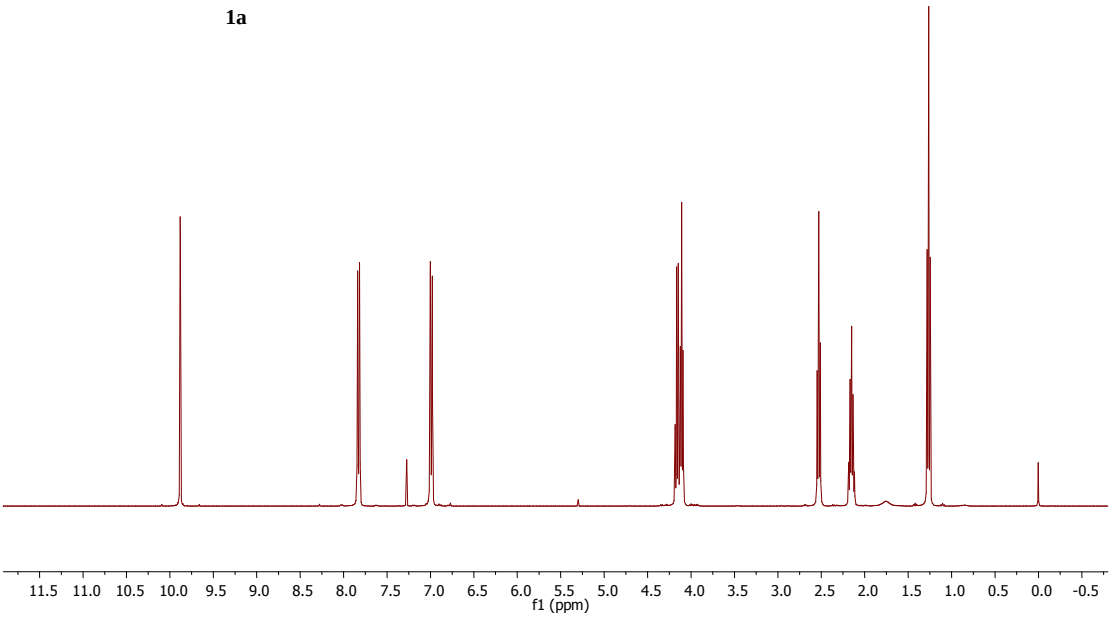


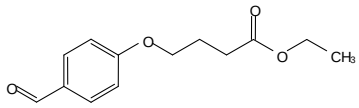
7k

Yield 73%, HPLC (Method B, t_R = 4.49 min), ^{1}H NMR (700 MHz, DMSO-d_6) δ 7.88 (d, J = 7.5 Hz, 2H), 7.71 (dd, J = 13.5, 7.3 Hz, 2H), 7.41 (t, J = 7.4 Hz, 2H), 7.32 (t, J = 14.8 Hz, 2H), 4.26 (m, 2H), 4.21 (t, 1H), 4.15 (t, J = 5.3 Hz, 1H), 4.08 (t, 1H), 3.01 (m, J = 6.7 Hz, 2H), 2.47 – 2.38 (m, 2H), 2.02 (s, 3H), 1.88 (m, 1H), 1.78 (m, J = 23.5, 15.3 Hz, 1H), 1.62 (m, 1H), 1.50 (m, 1H), 1.37 (m, 2H), 1.22 (m, 2H). ^{13}C NMR (175 MHz, DMSO-d_6) δ 173.6, 171.0, 160.8, 155.9, 143.7, 140.7, 127.6, 127.0, 125.3, 120.1, 65.6, 53.9, 52.1, 46.6, 36.9, 31.7, 29.6, 28.6, 22.5, 14.6. MS (ESI) $[MH]^+$ calcd. $C_{27}H_{34}N_4O_5S$ 526.6, found 526.0.

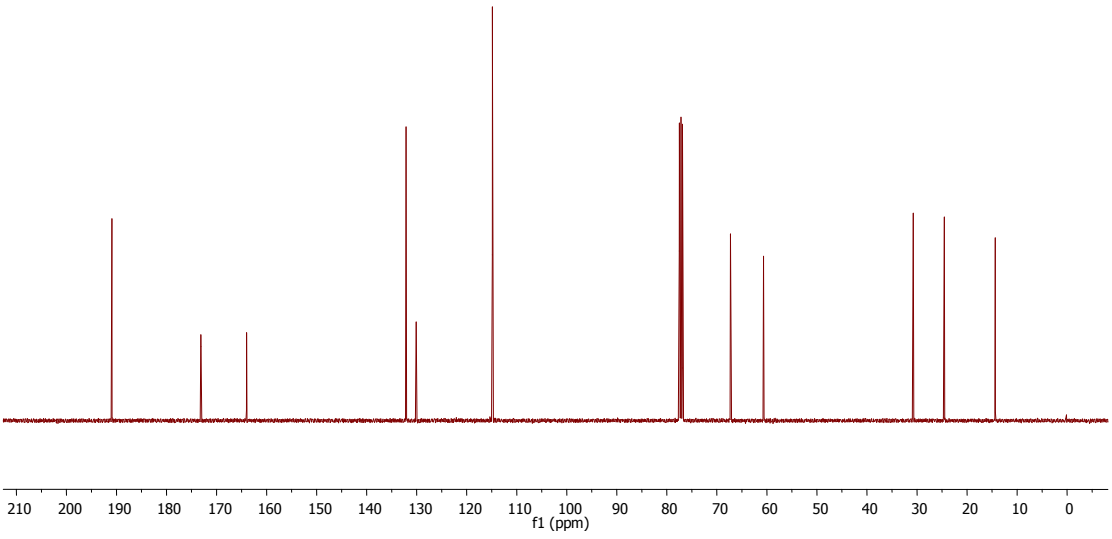


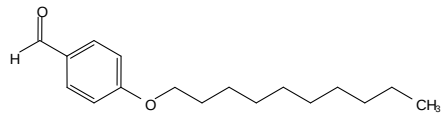
1a


11.5 11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)

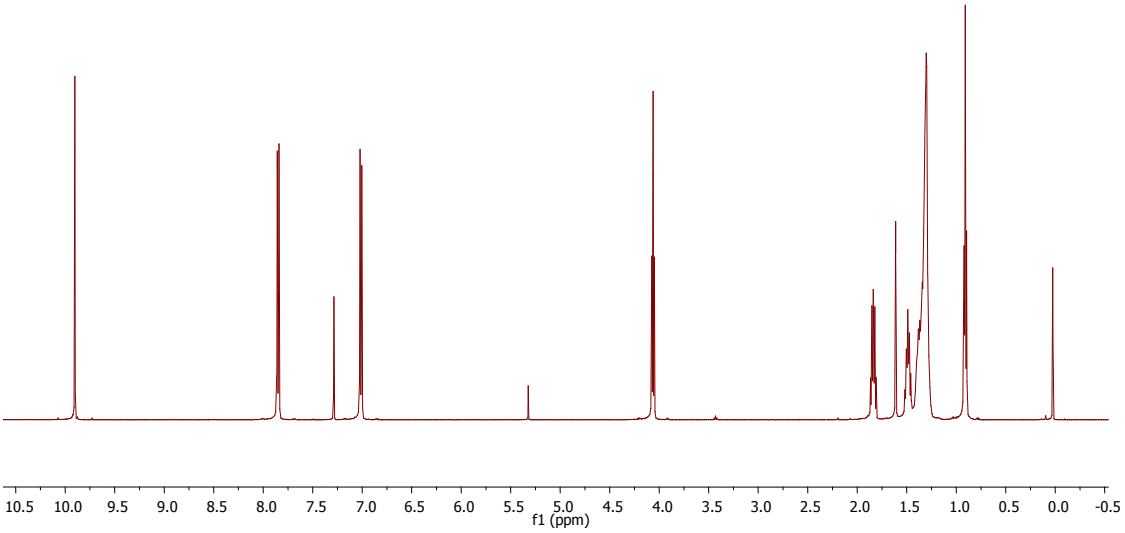


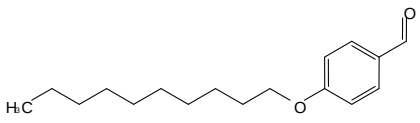
1a

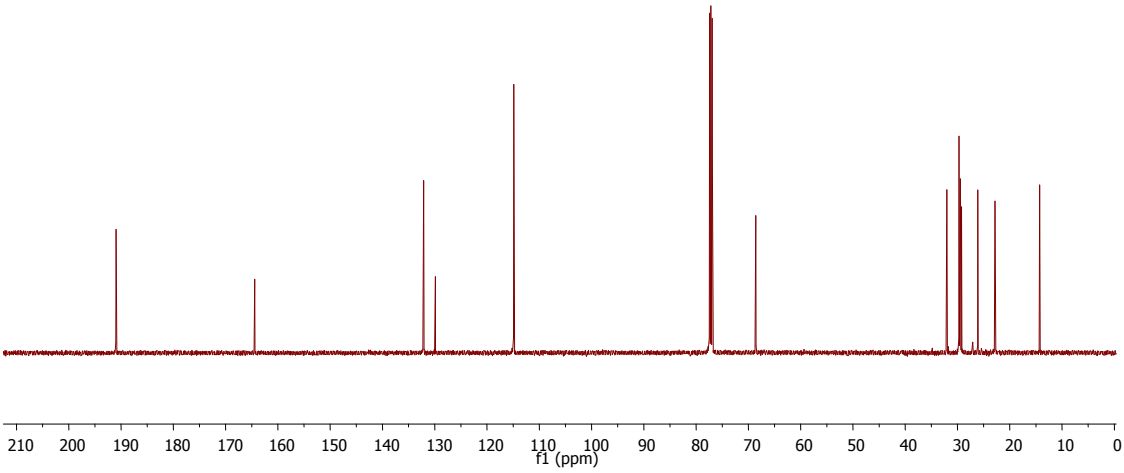

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

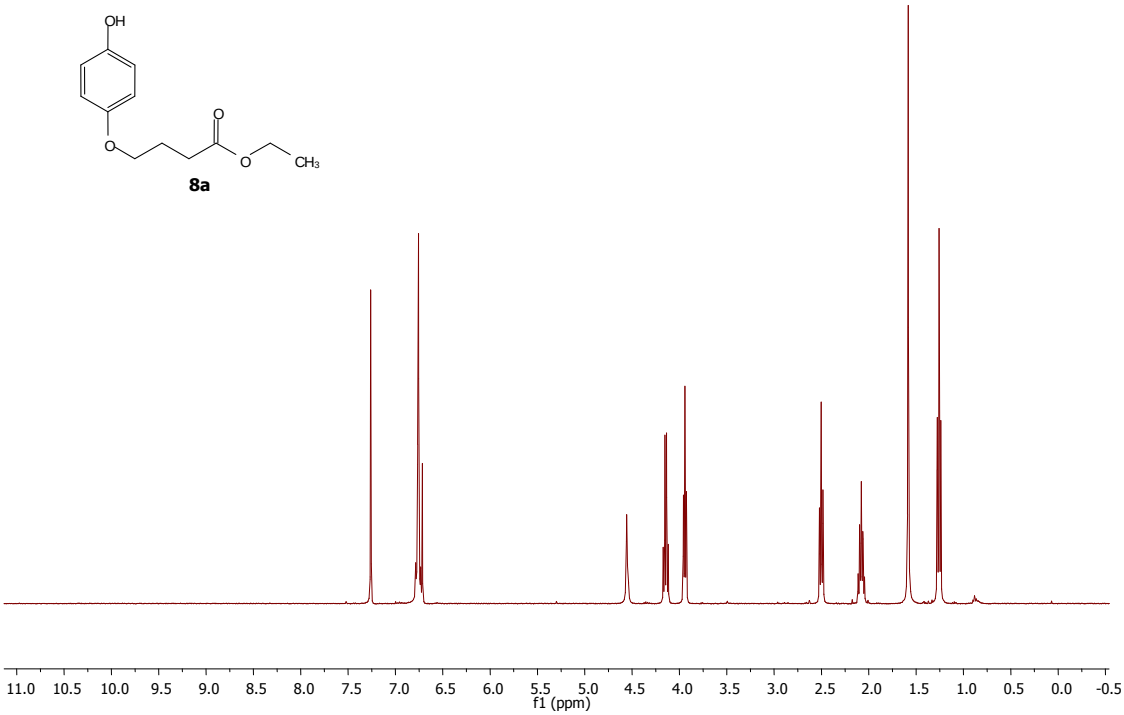

O
H
O
CH3

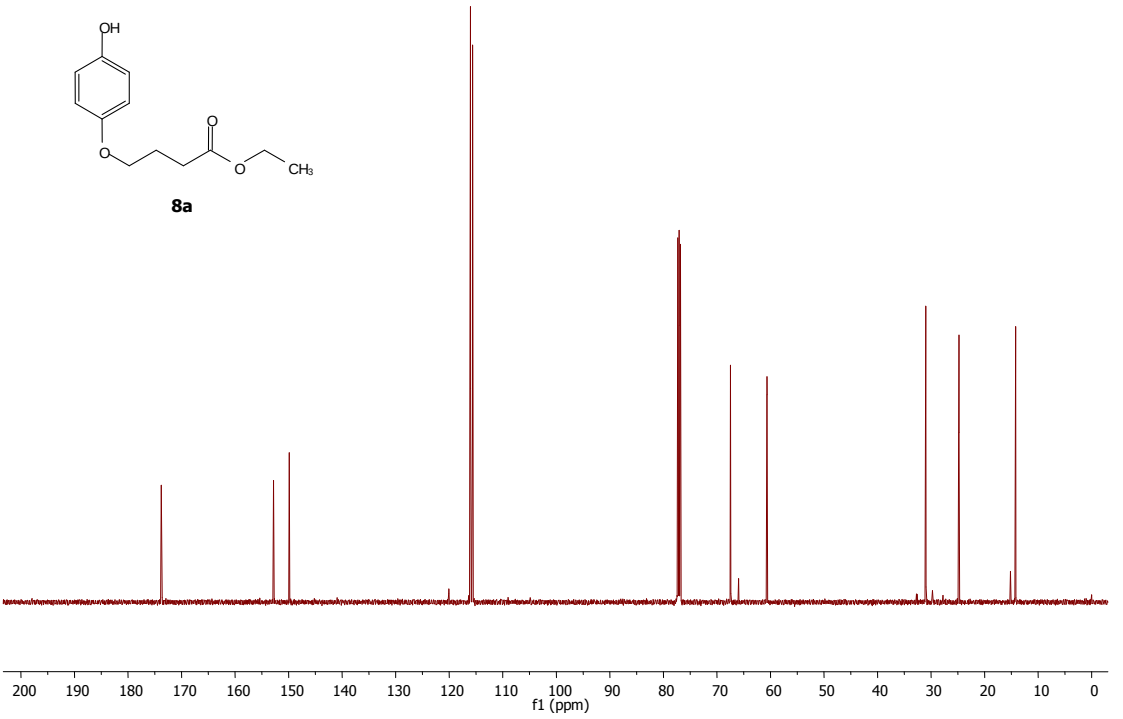
1b

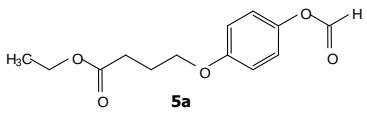

10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)


O
H3C
O

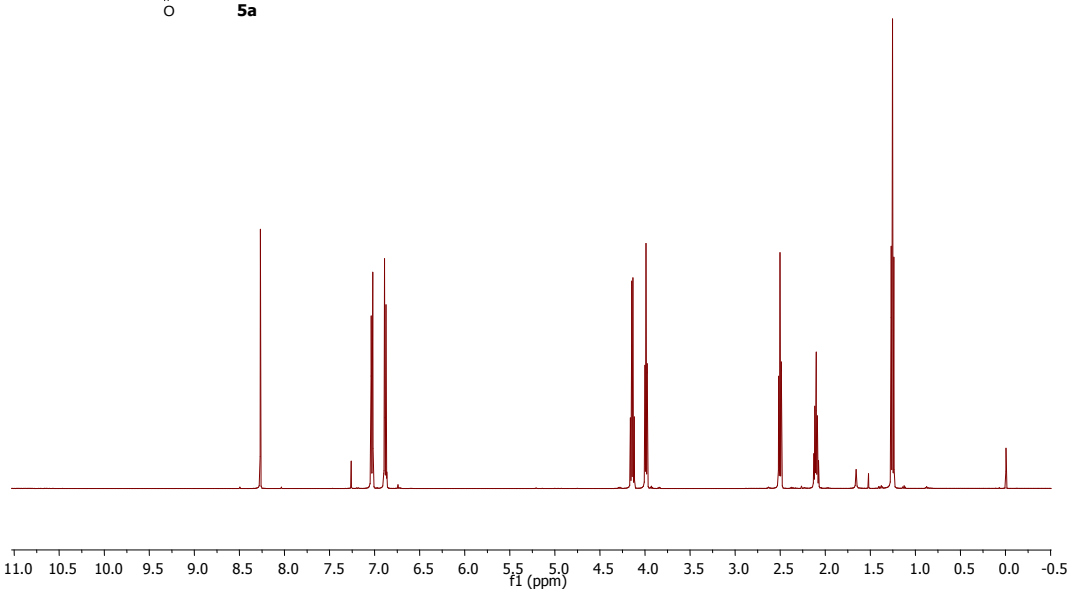

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

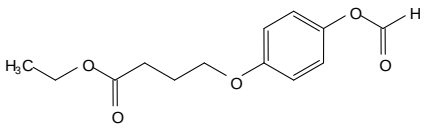

OH
O
O
O
CH3
8a
11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)


OH
O
O
O
CH3
8a
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

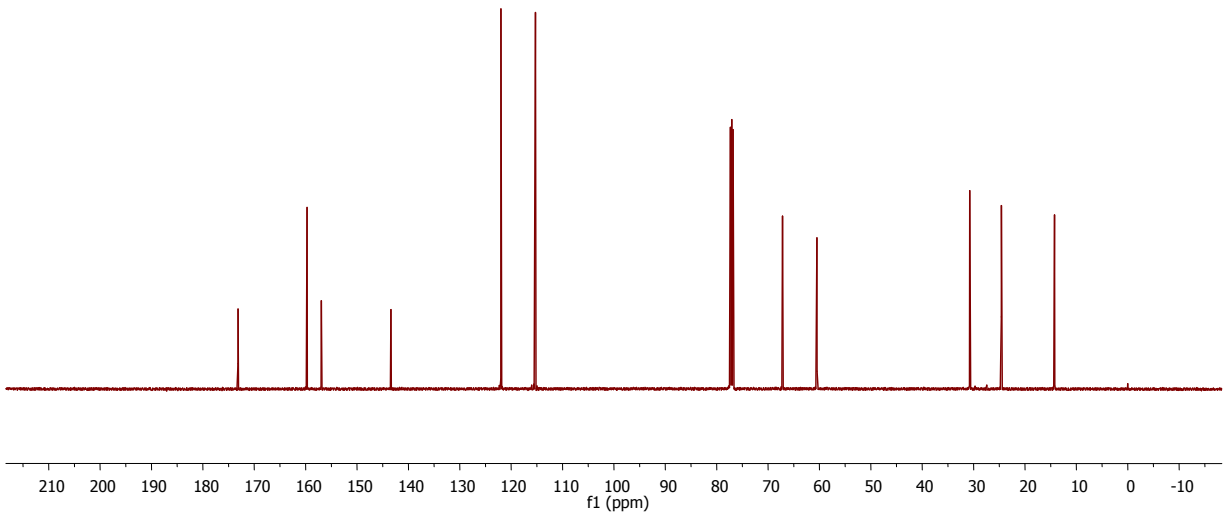

H3C
O
O
O
O
O
H

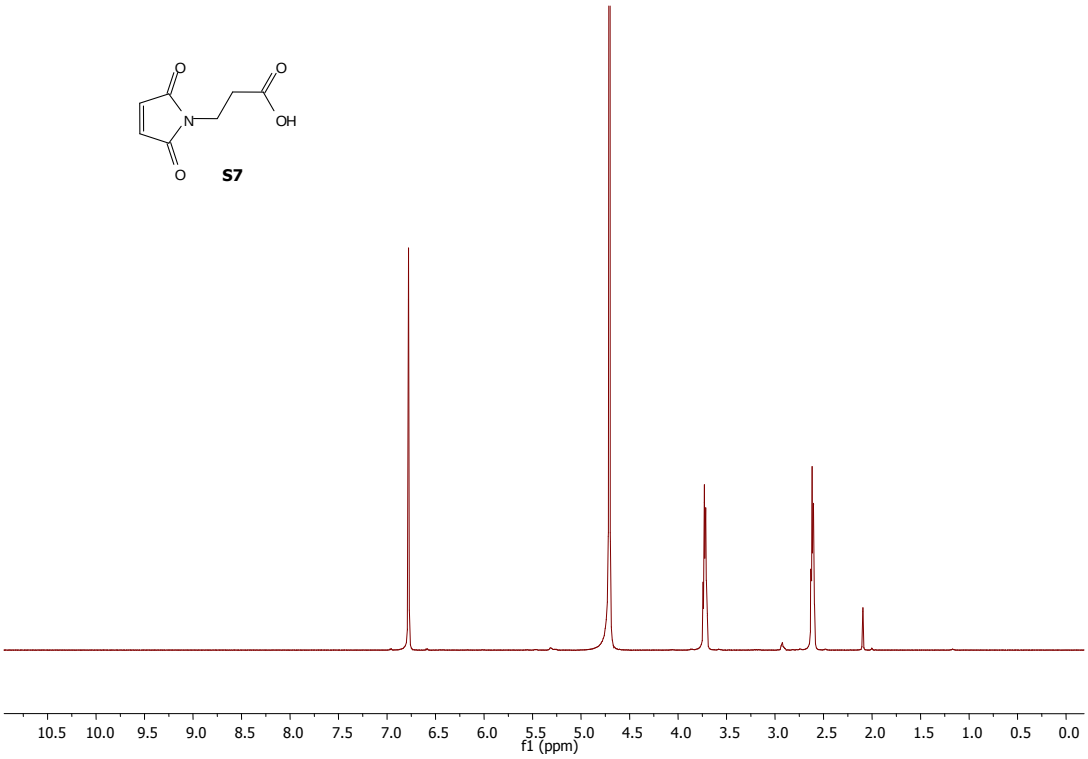
5a

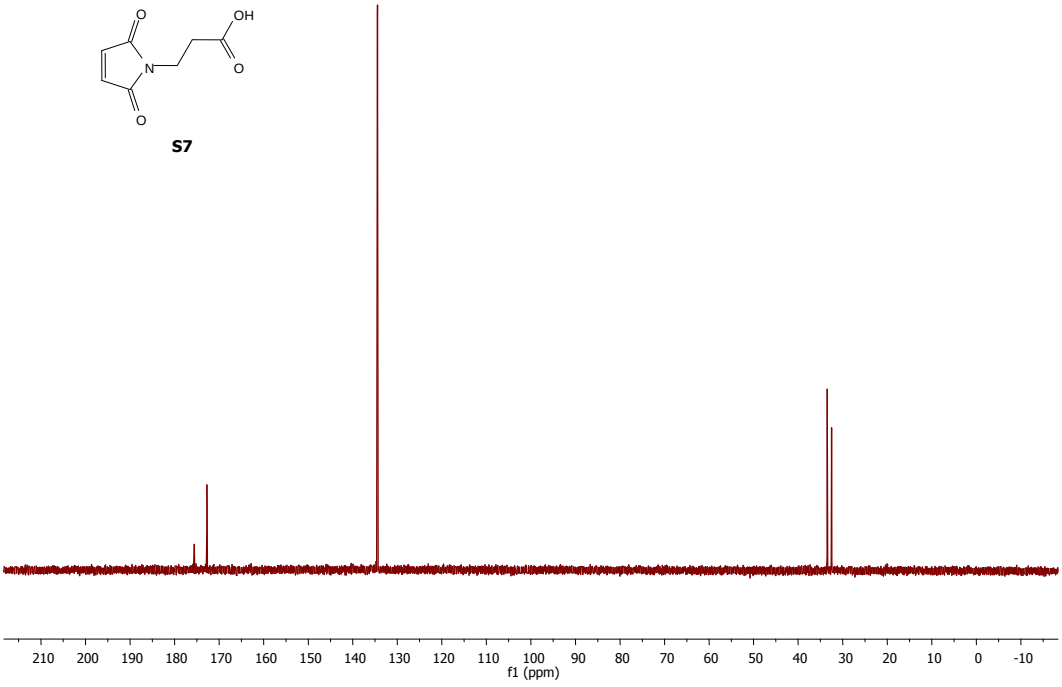

11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)

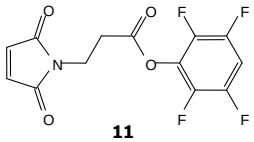

H3C
O
O
O
O
O
H

5a


210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

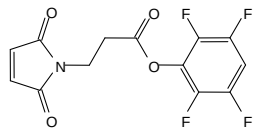

O
O
N
OH
O
S7
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)


O
OH
N
O
O
S7
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)


O
O
F
F
N
O
O
F
F

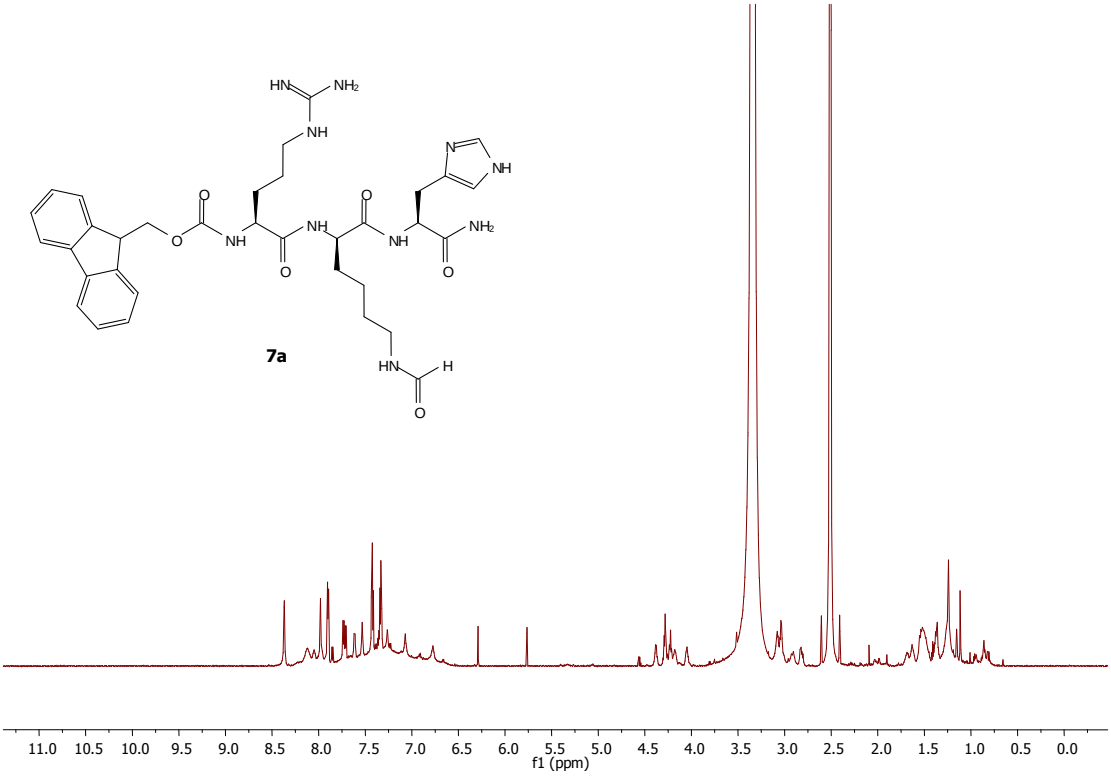
11

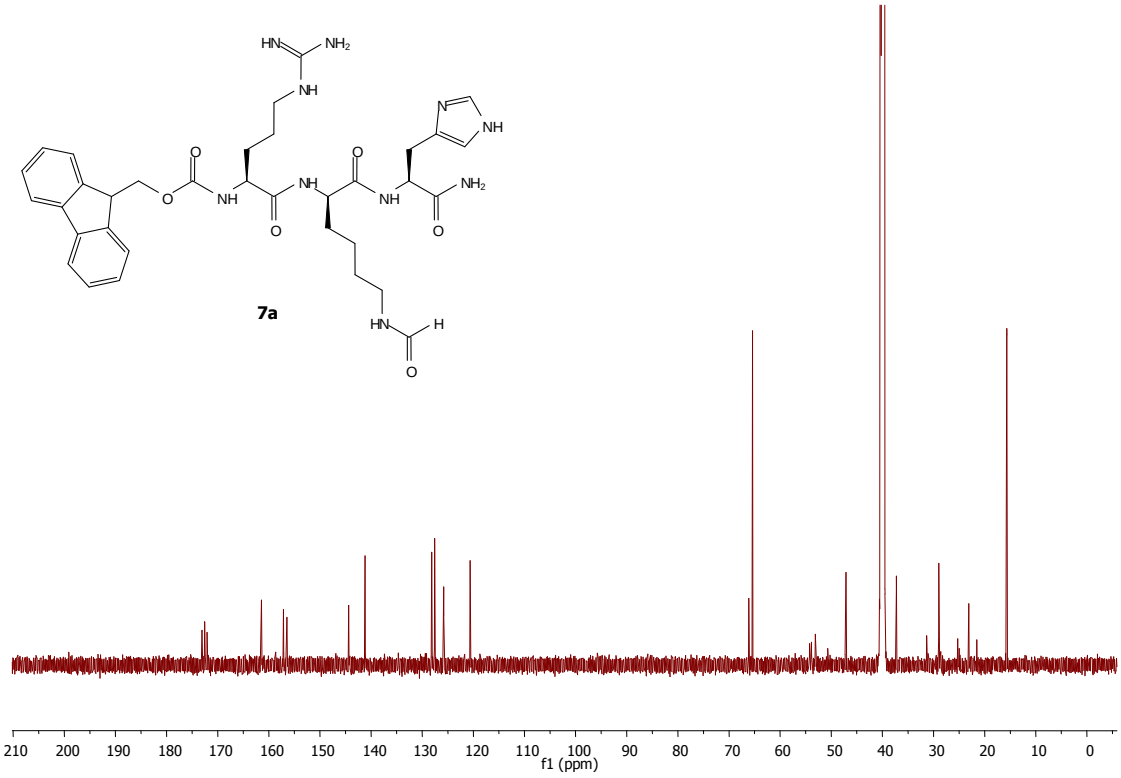
10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1.0
f1 (ppm)

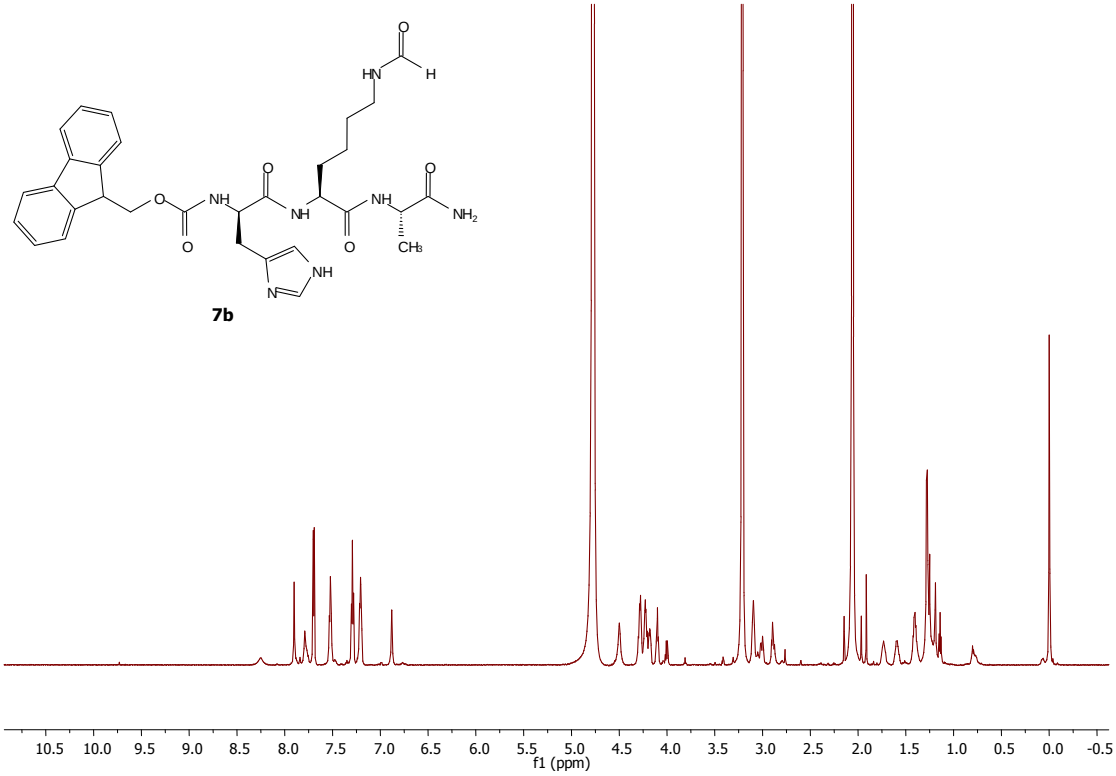

O
O
F
F
N
O
O
F
F

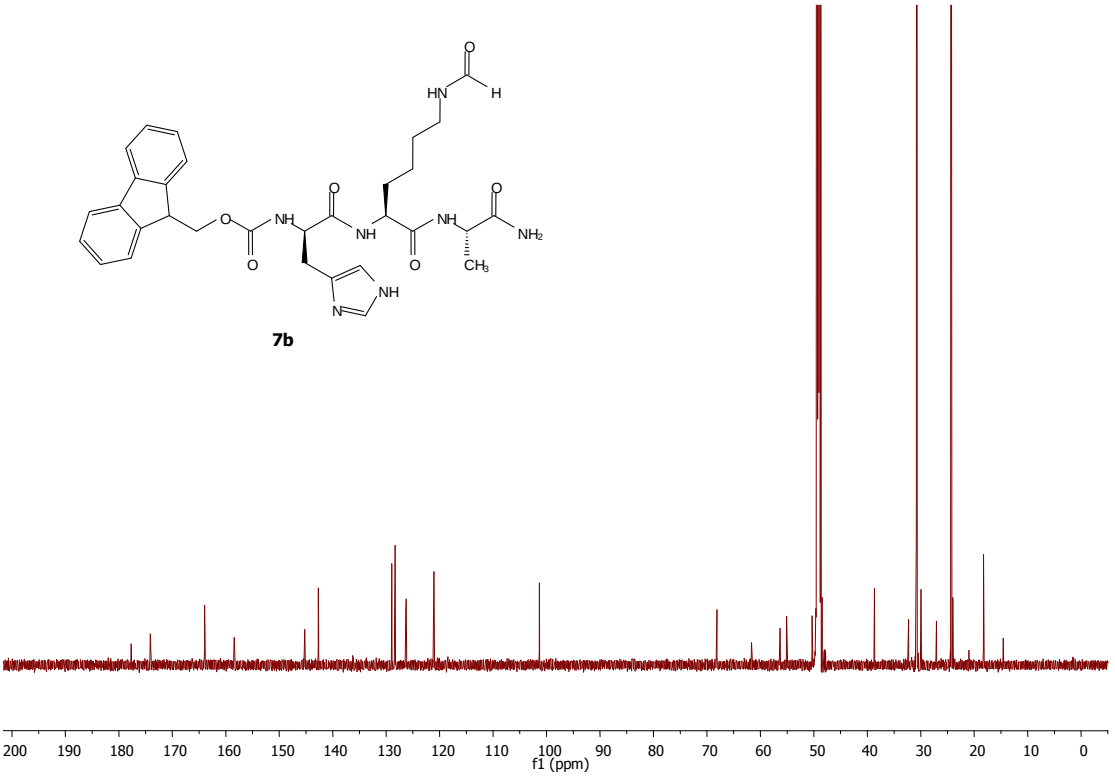
11

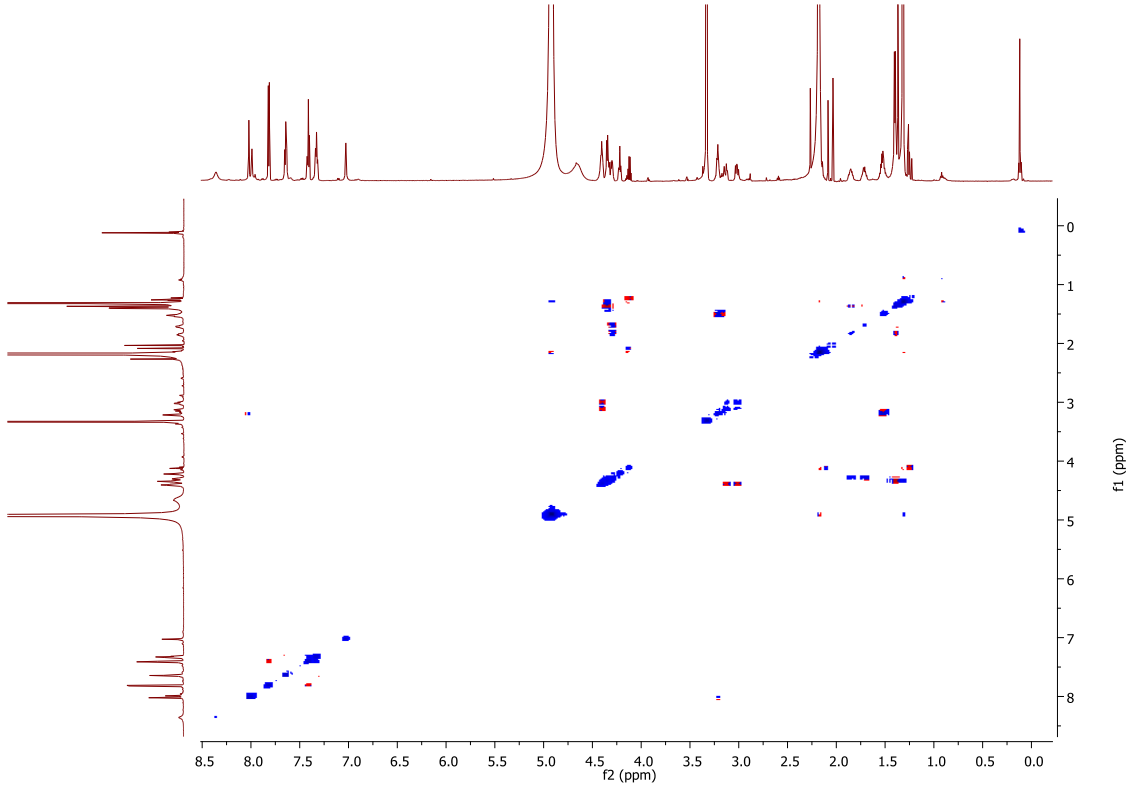
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

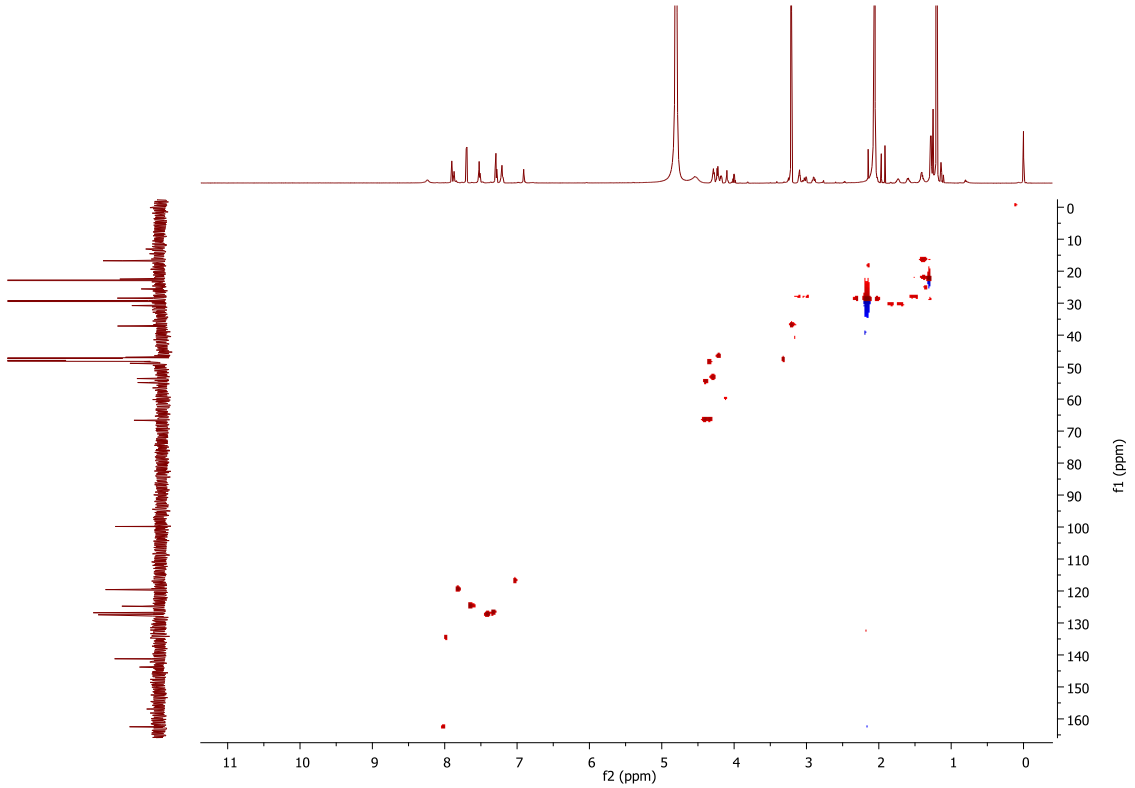

HN
NH2
NH
N
NH
O
NH
NH
NH
NH2
O
O
O
O
7a
HN
H
O
11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

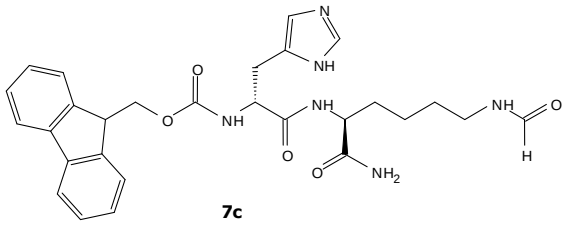

HN
NH2
NH
N
NH
O
NH
NH
NH
NH2
O
O
O
O
7a
HN
H
O
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)


7b
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)


7b
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

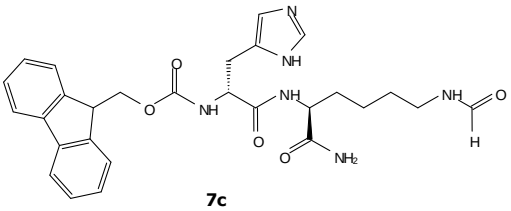

f2 (ppm)
f1 (ppm)


f2 (ppm)
f1 (ppm)


N
NH
O
O
NH
NH
O
NH
O
H
O
NH2

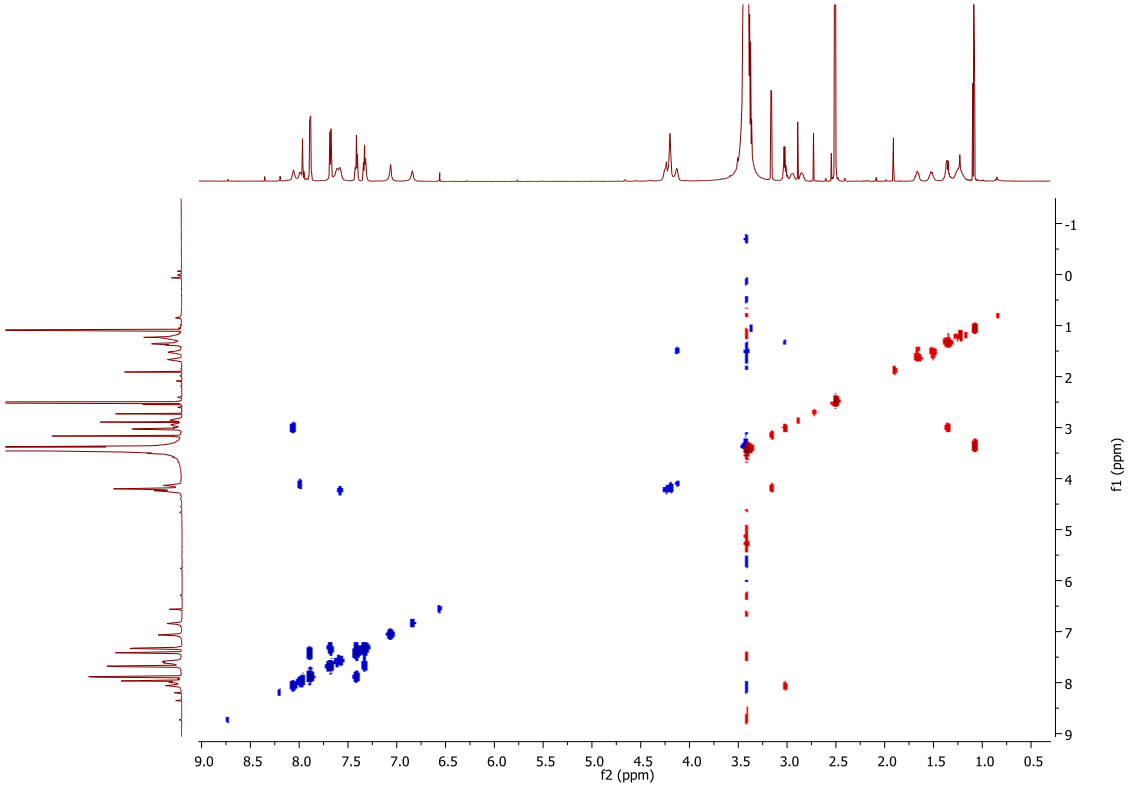
7c

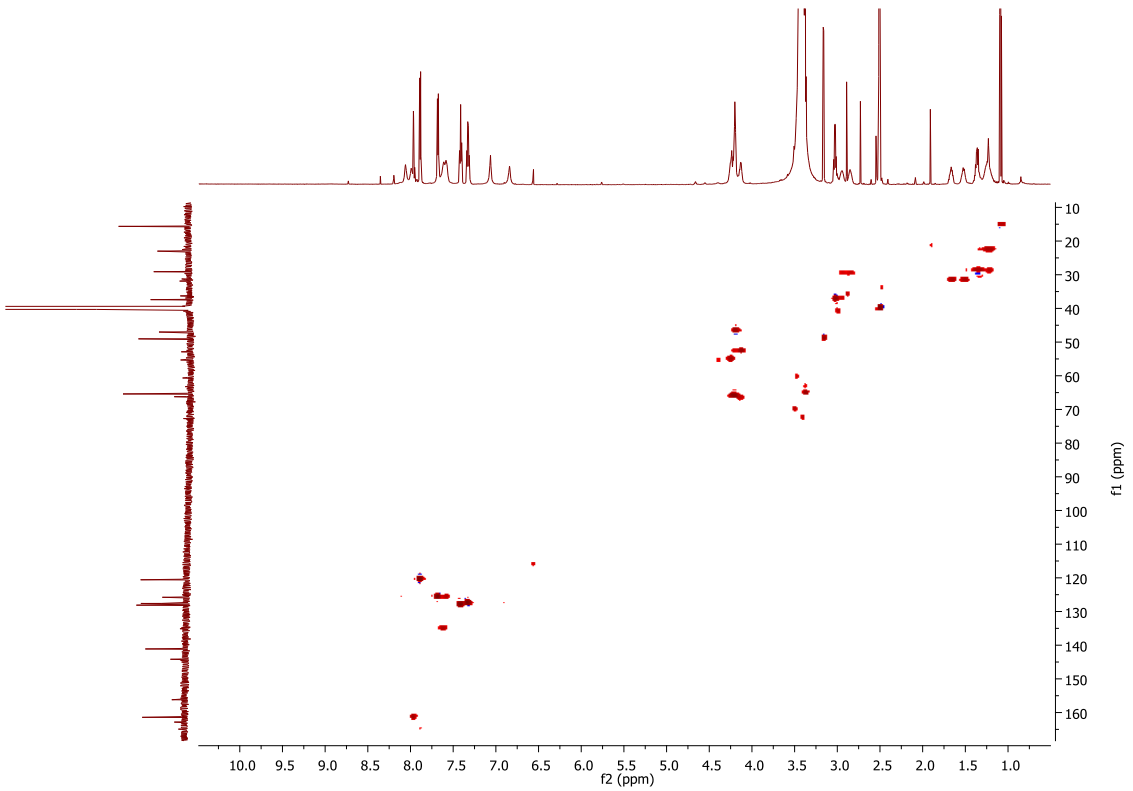
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

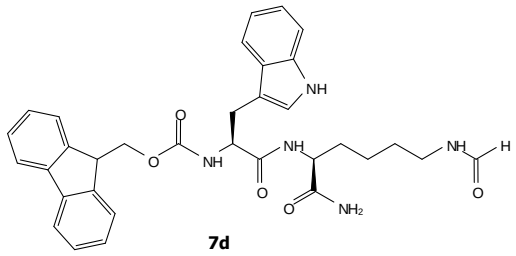

N
NH
O
O
NH
NH
O
NH
O
H
O
NH2

7c

200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

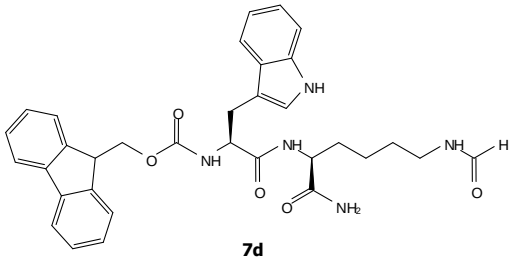

f2 (ppm)
f1 (ppm)


f2 (ppm)
f1 (ppm)


NH
O
O
NH
NH
O
O
NH
NH
H
O
NH2

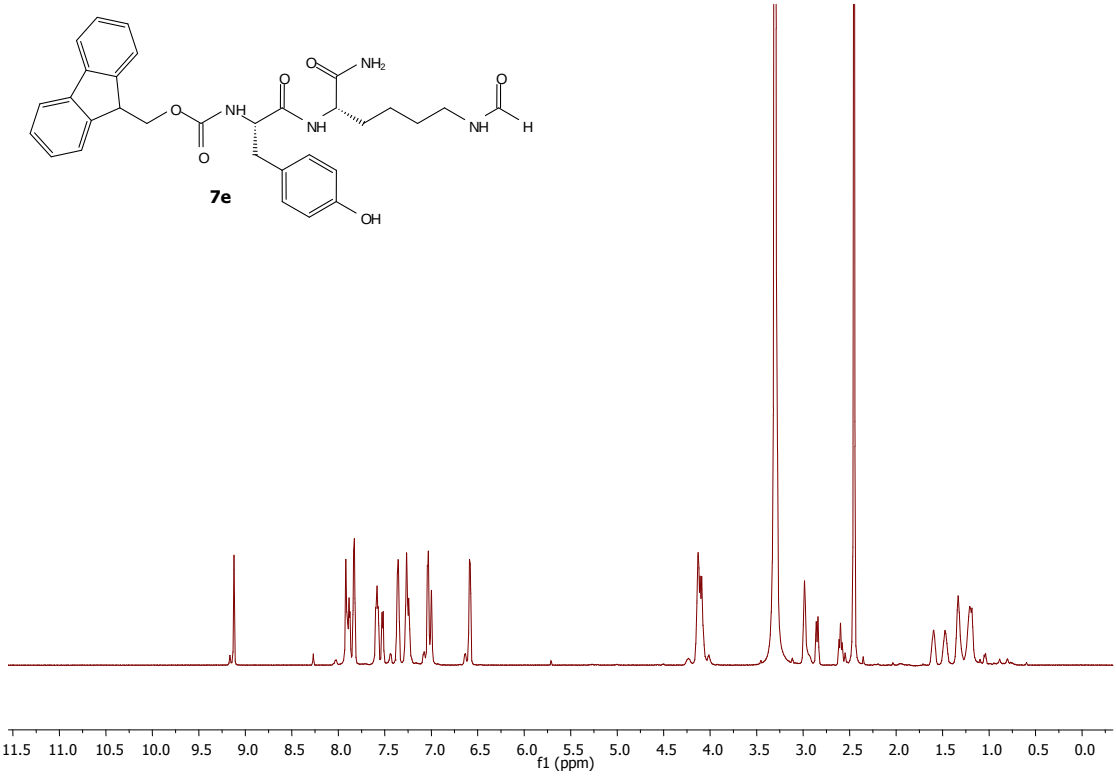
7d

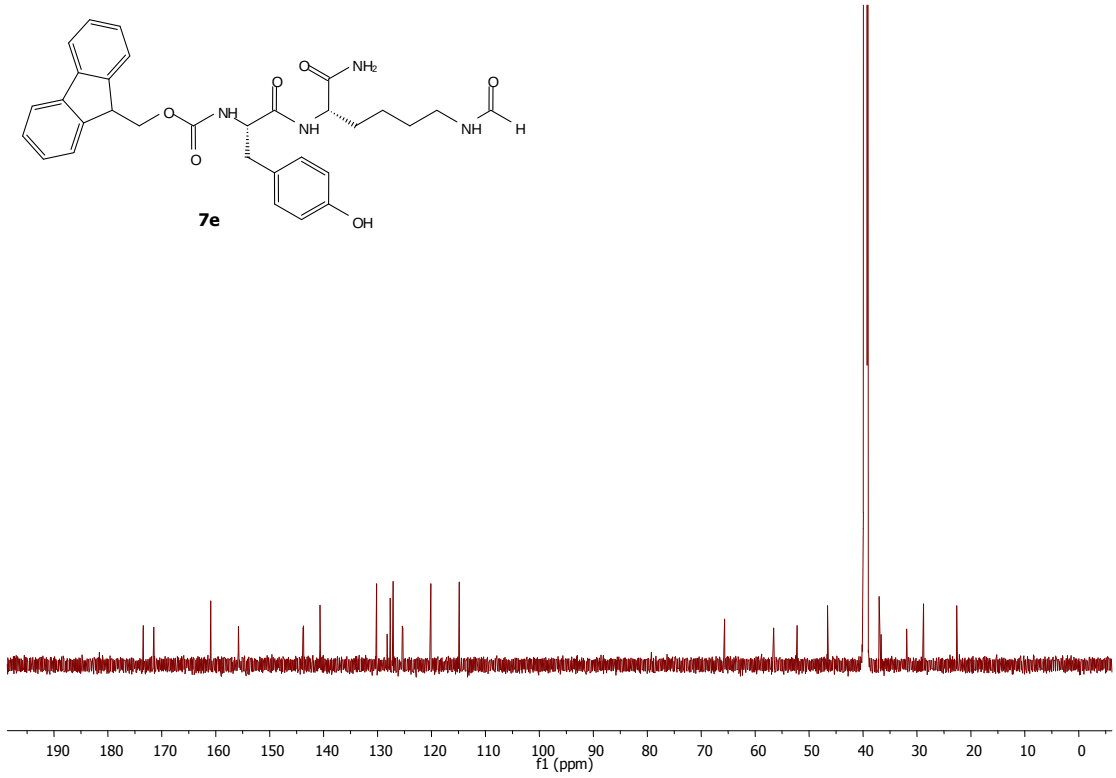
10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

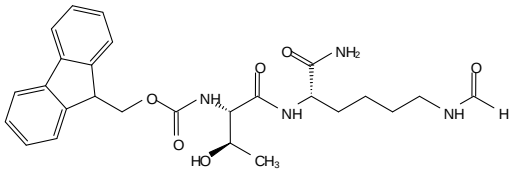

NH
O
O
NH
NH
O
O
NH
NH
H
O
NH2

7d

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

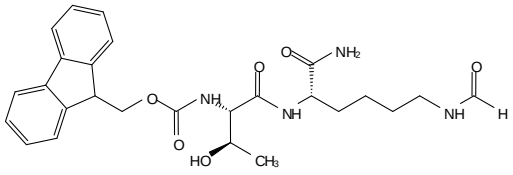

O
NH
O
NH
O
NH2
NH
H
O
O
OH
7e
11.5 11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)


O
NH
O
NH
O
NH2
NH
H
O
O
OH
7e
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)


O
NH2
O
NH
O
NH
NH
H
O
HO
CH3

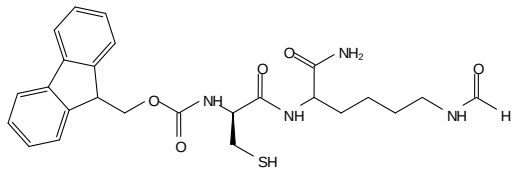
7f

10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)

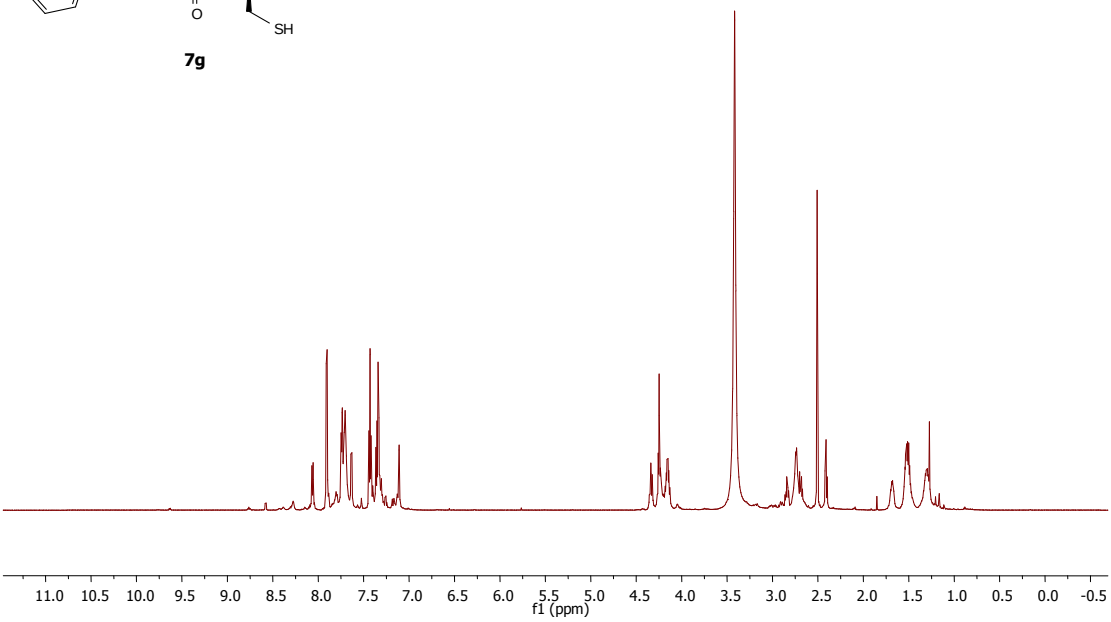

O
NH2
O
NH
O
NH
NH
H
O
HO
CH3

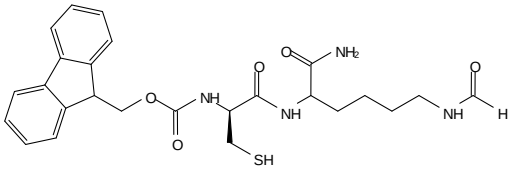
7f

200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)


O
NH2
O
NH
O
NH
NH
O
H
O
SH

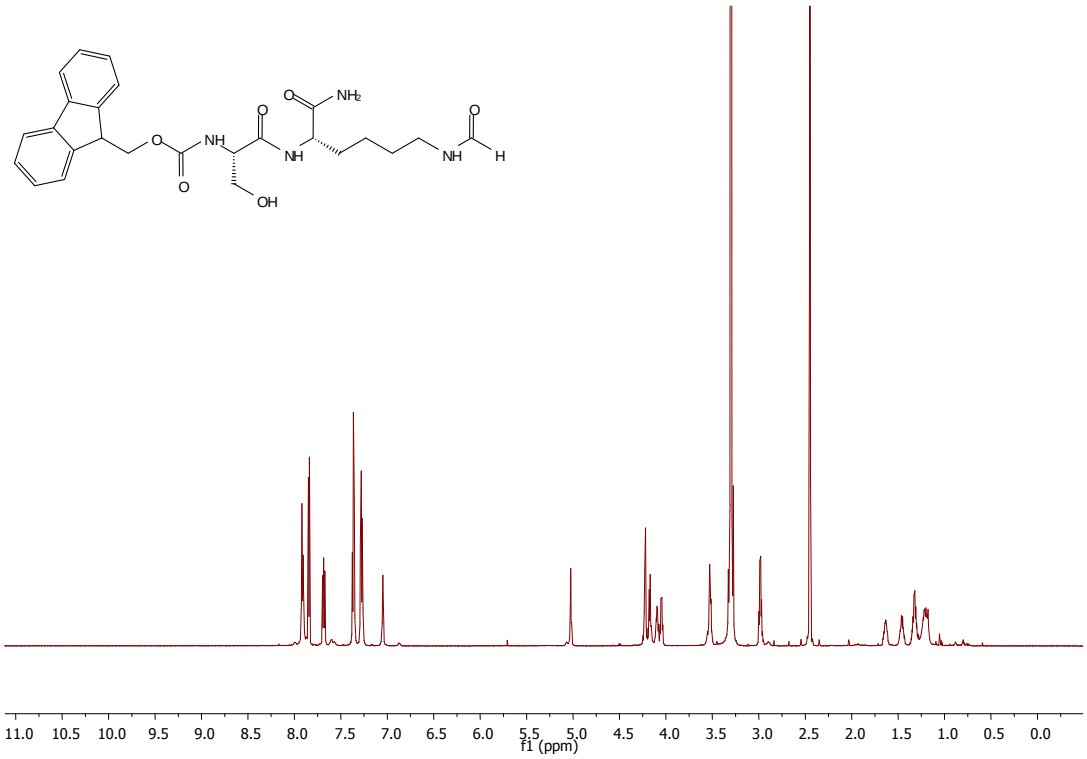
7g

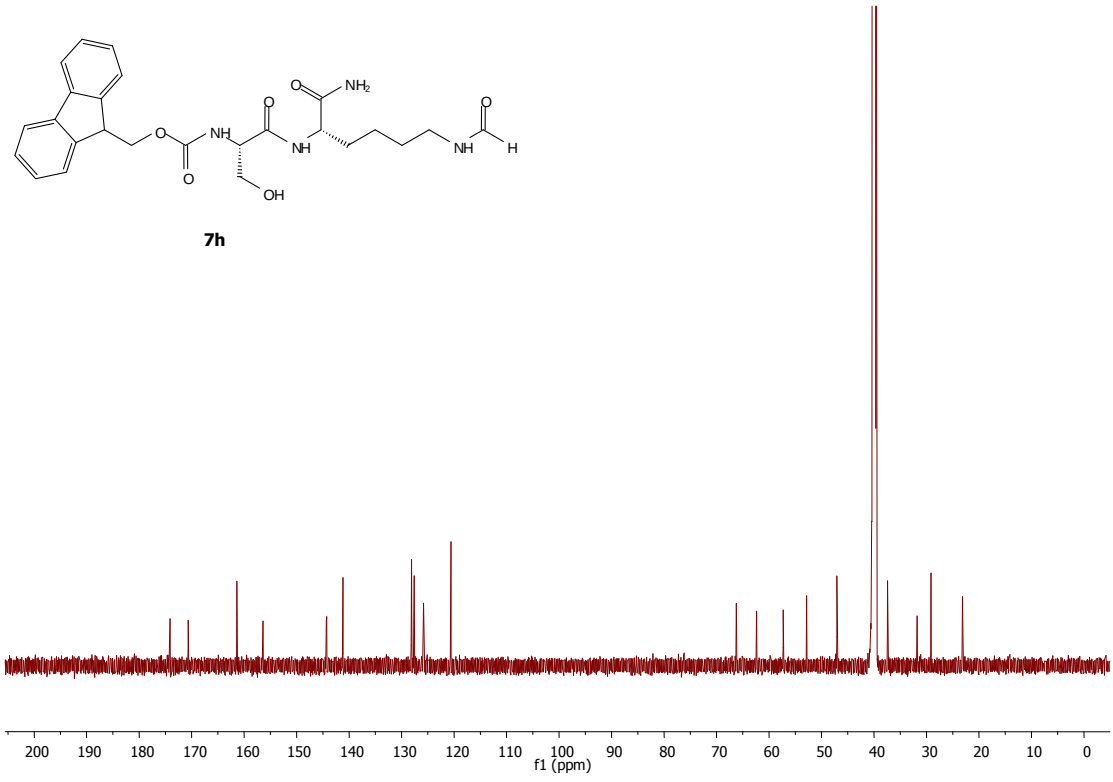

11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)


O
NH2
O
NH
O
NH
NH
O
H
O
SH

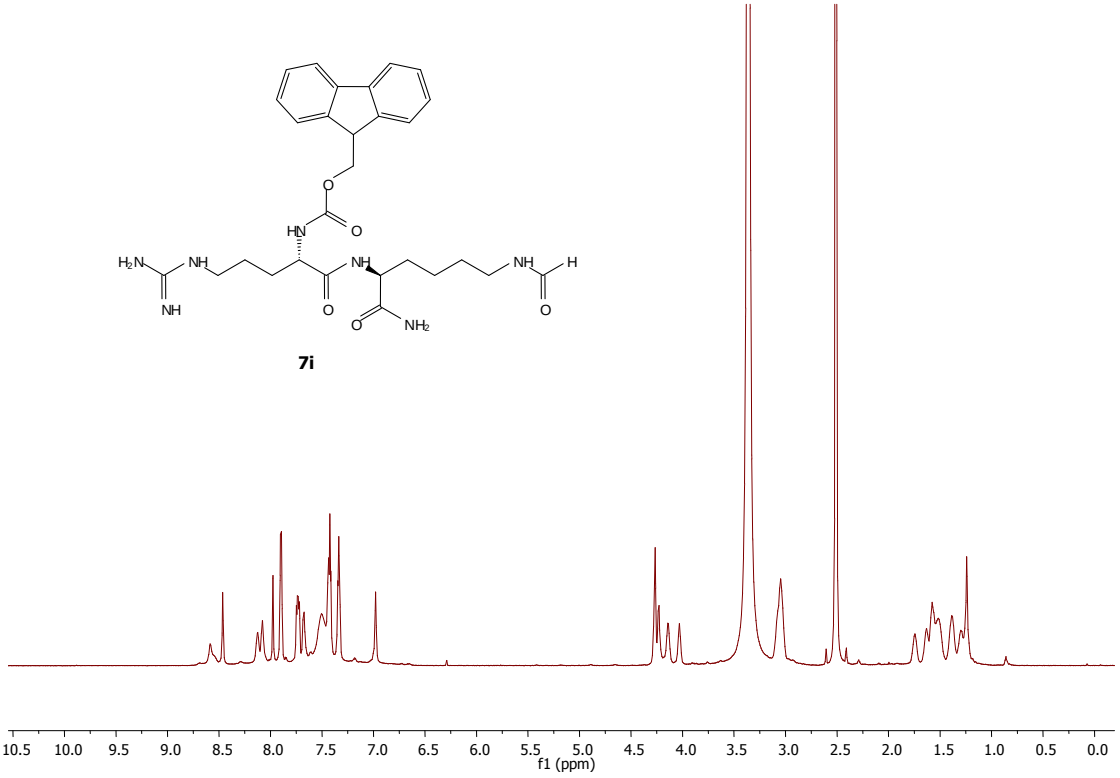
7g

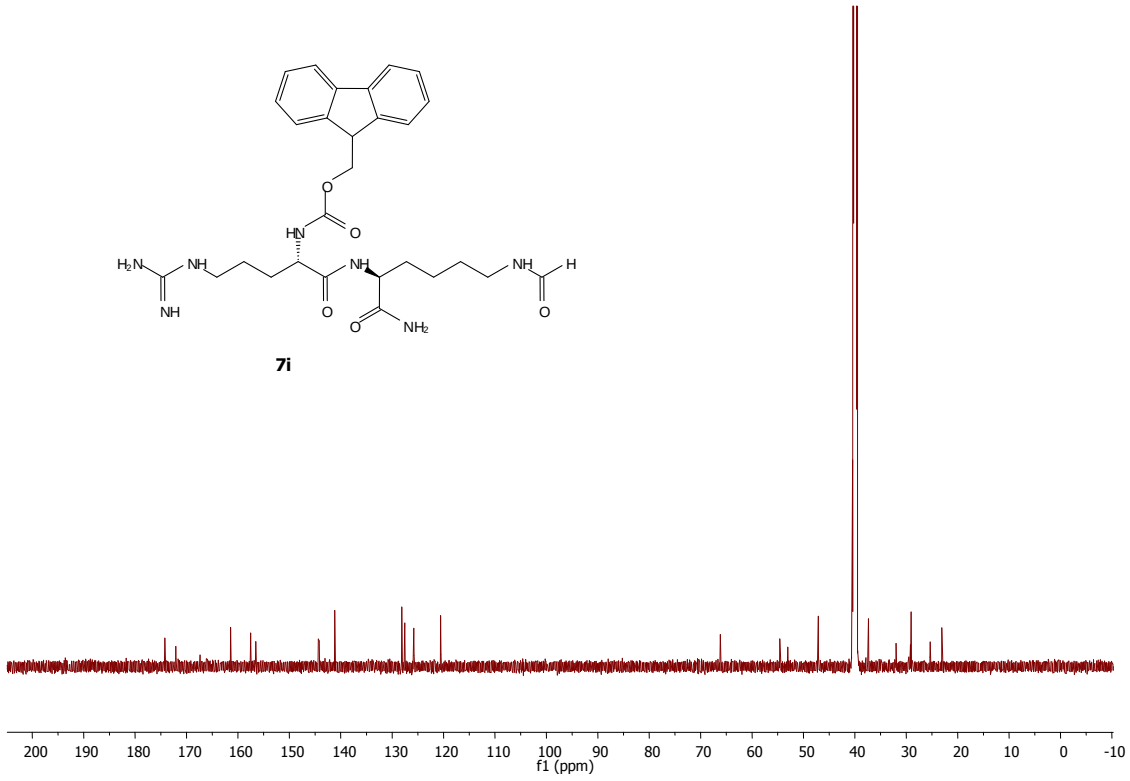
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

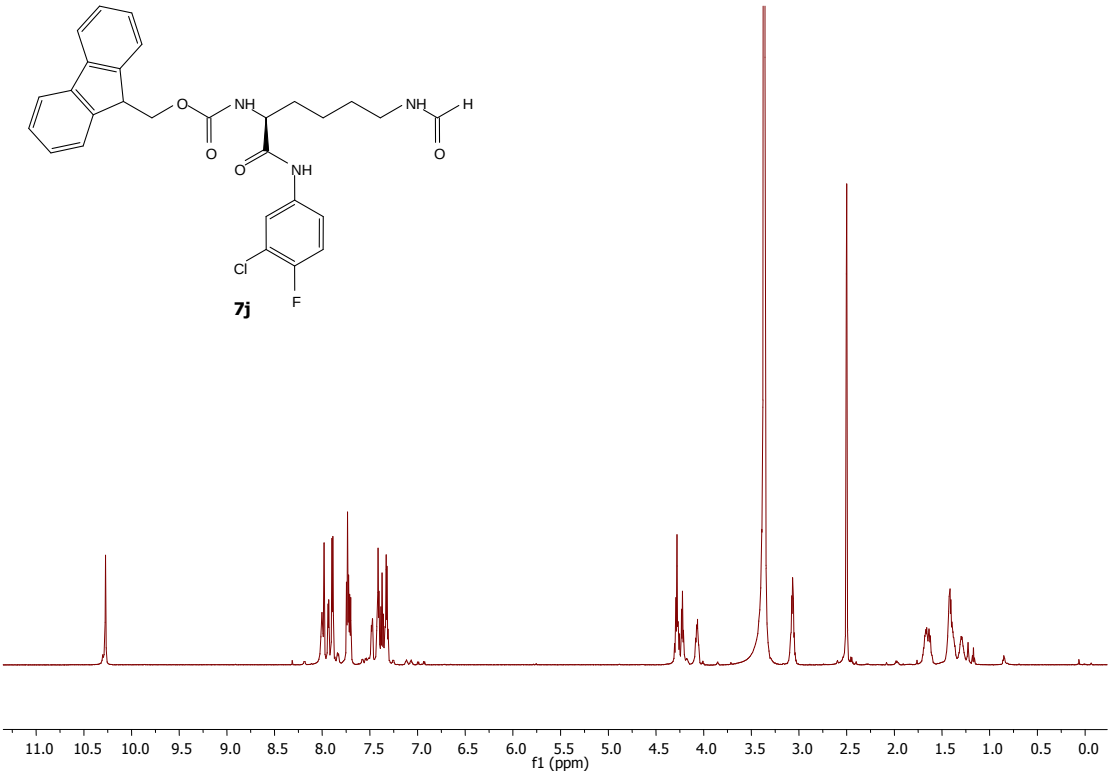

O
NH
O
NH
O
NH2
NH
O
H
O
OH
11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)


O
NH
O
NH
O
NH2
NH
O
H
O
OH
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

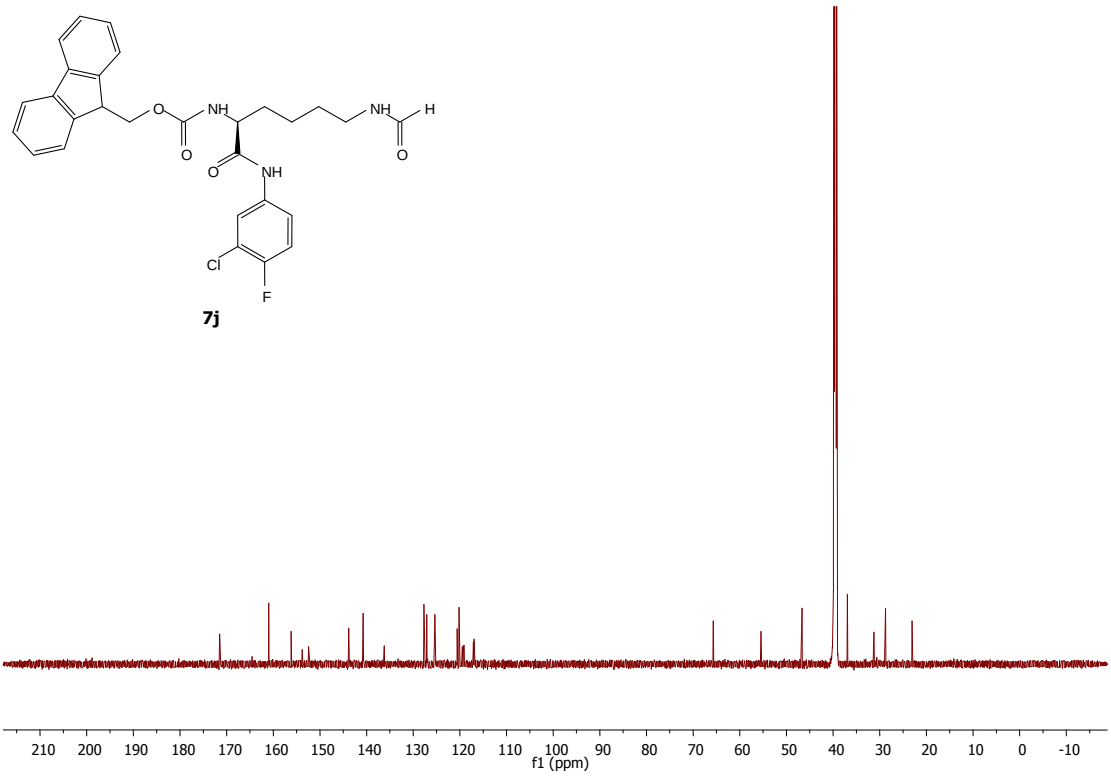
7h


7i
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

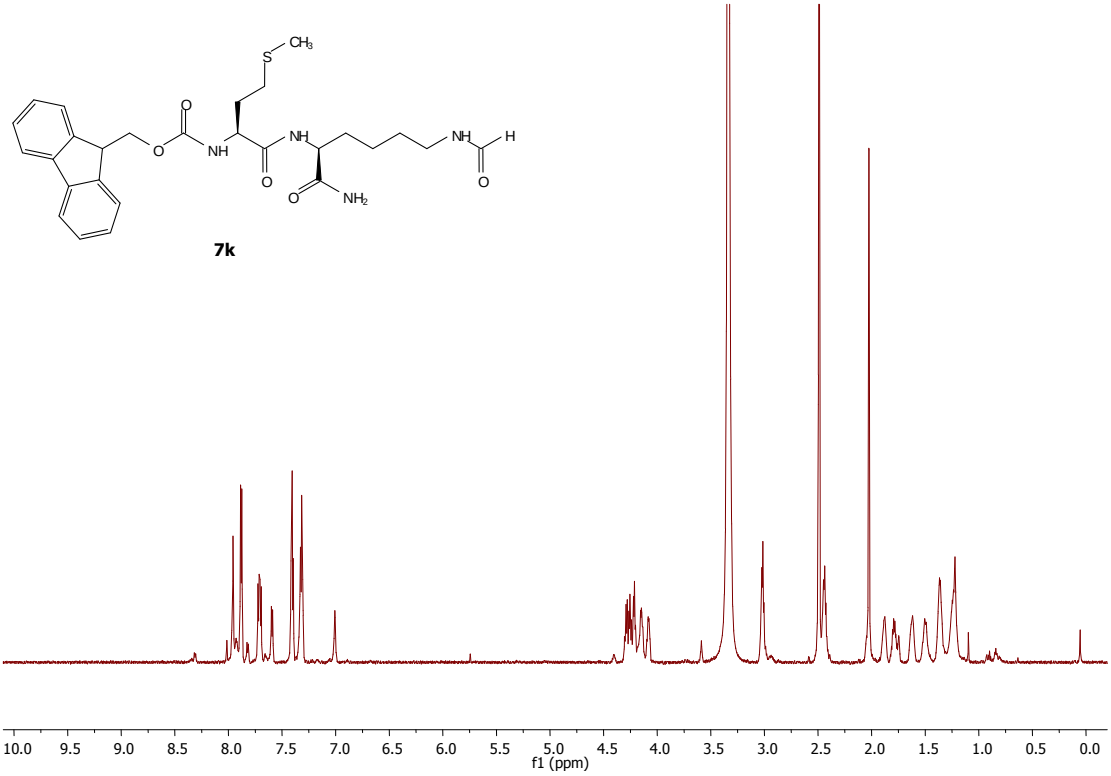

7i
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)


O
NH
NH
H
O
O
O
NH
Cl
F
11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

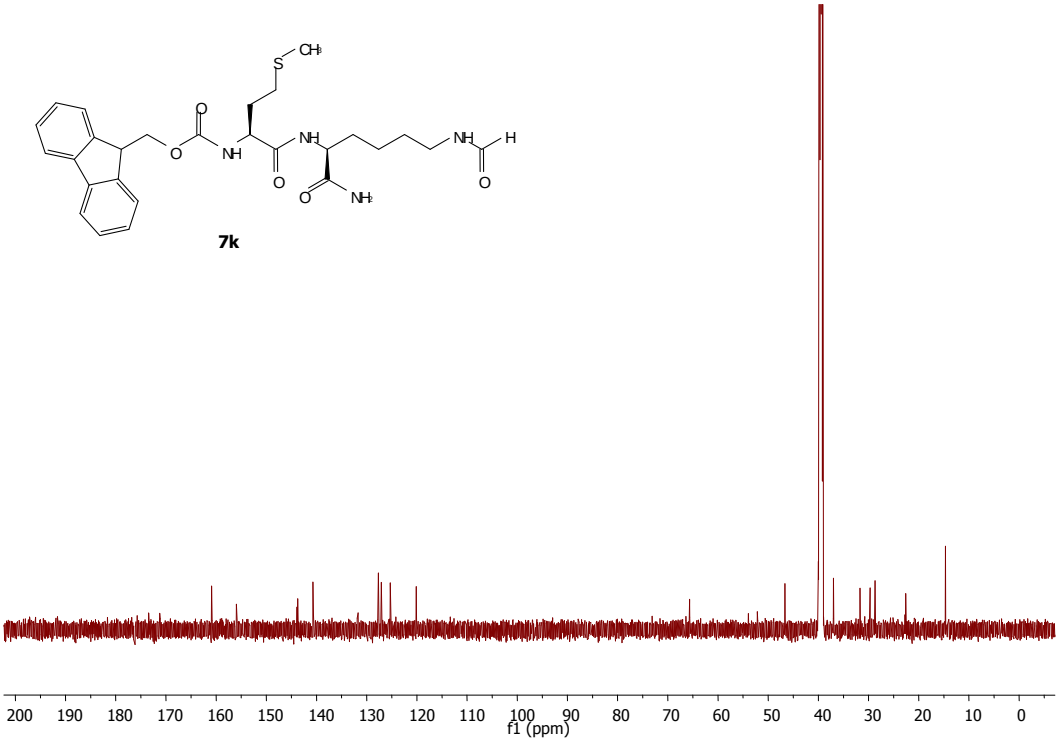
7j


O
NH
NH
H
O
O
O
NH
Cl
F
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

7j


S
CH3
O
O
NH
NH
O
NH
H
O
O
NH2
10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

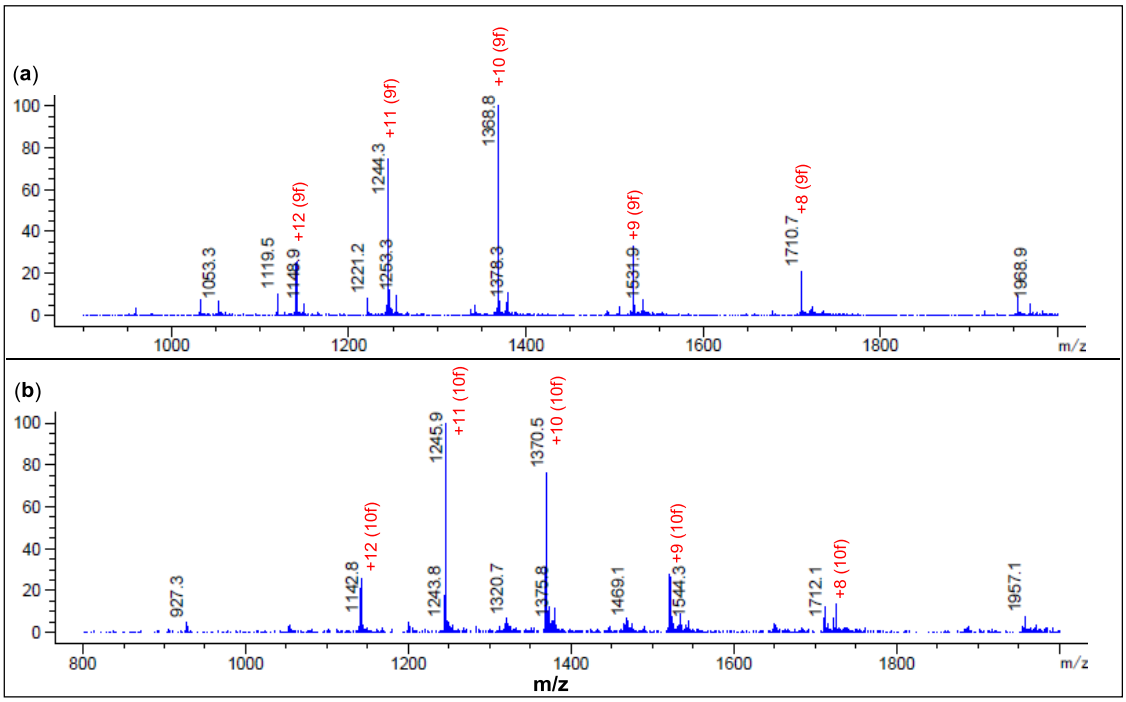
7k


S
CH3
O
O
NH
NH
O
NH
H
O
O
NH2
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

7k

7. MS and MS-MS

7.1. MS and MS-MS spectra for formylation of proteins using aldehyde **1** (*refer Figure 5*)



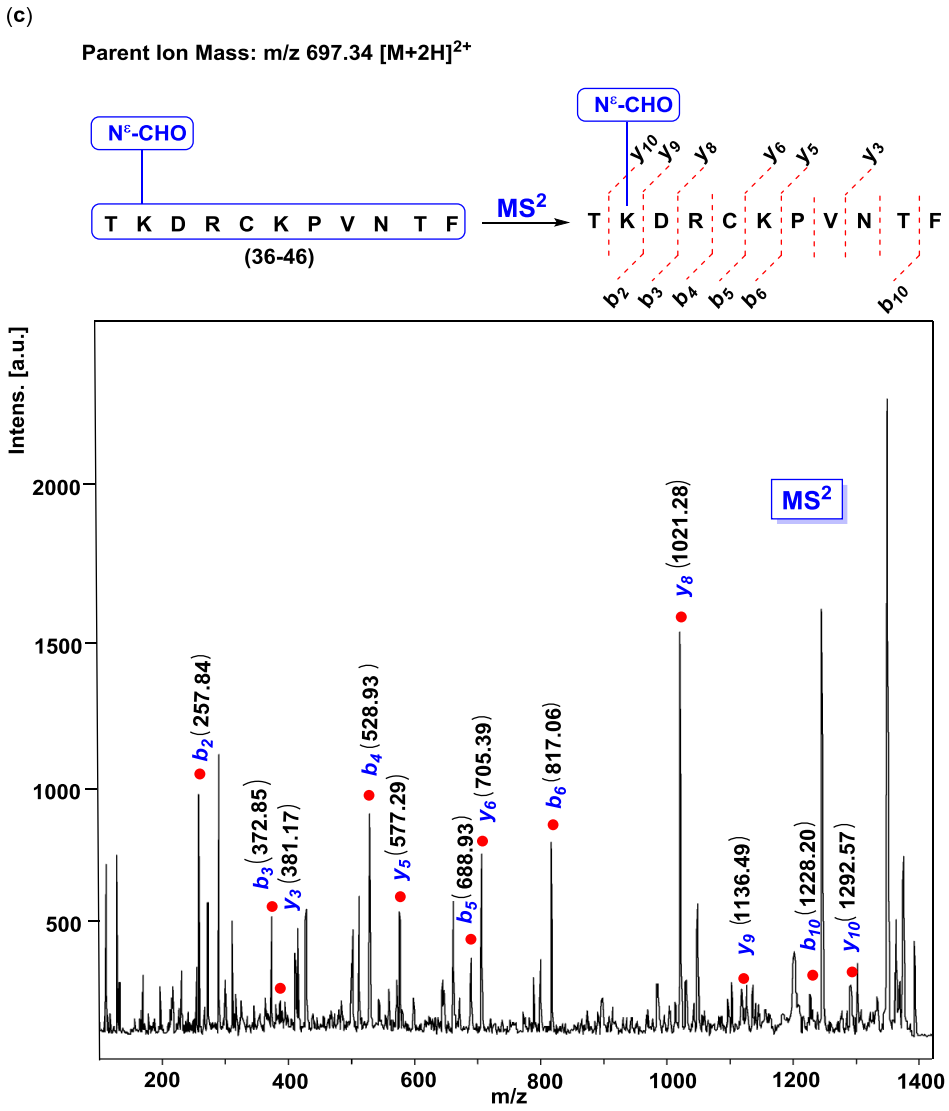


Figure S11. (a) ESI-MS spectra of native RNase A **9a** (MW 13681 Da). (b) ESI-MS spectra of RNase A (1 equiv.) and aldehyde (600 equiv.): mono-formylated product **10f**. (c) MS-MS spectrum of TKDRCKPVNTF (36-46). The site of modification is K37 in mono-formylated RNase A **10f**.

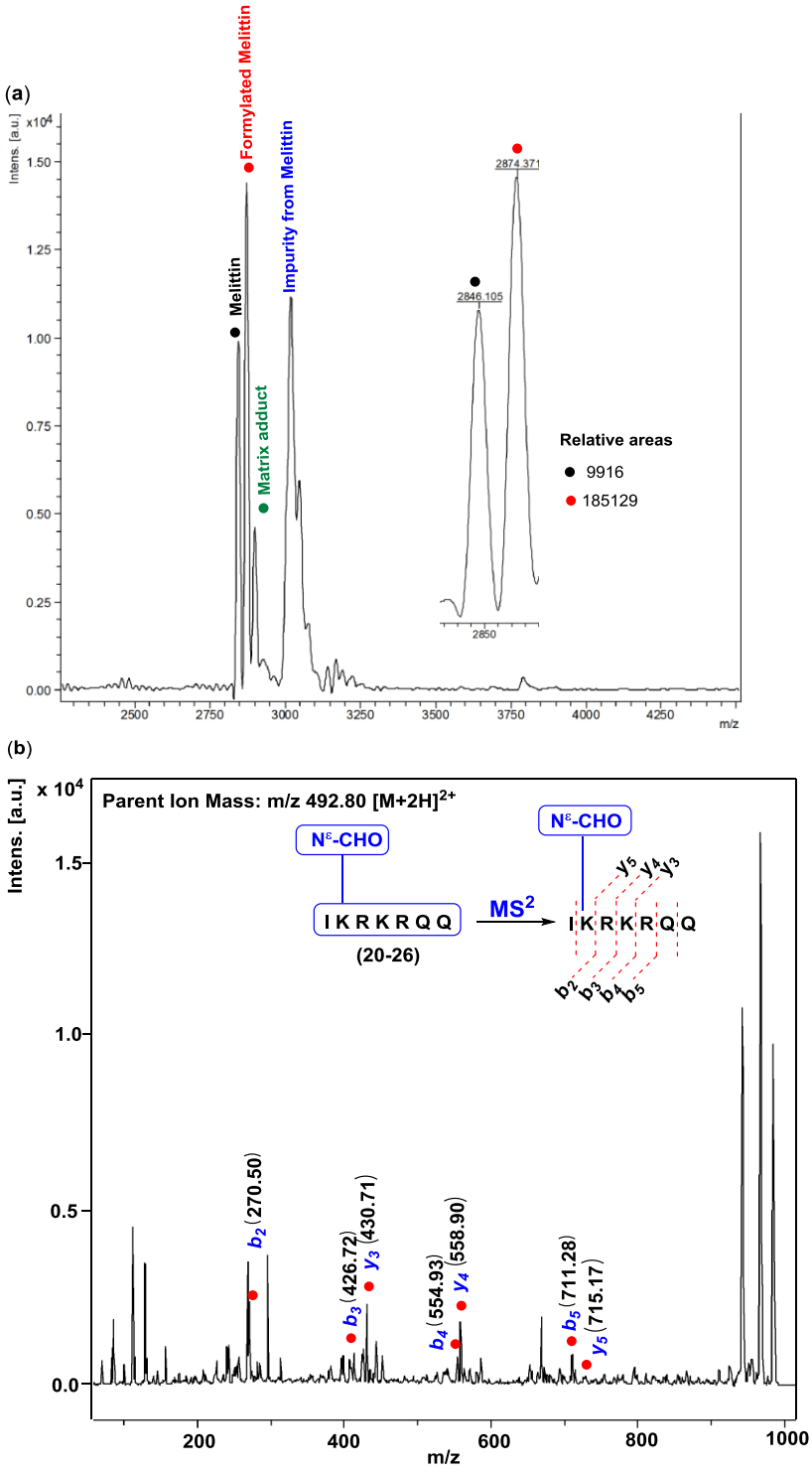
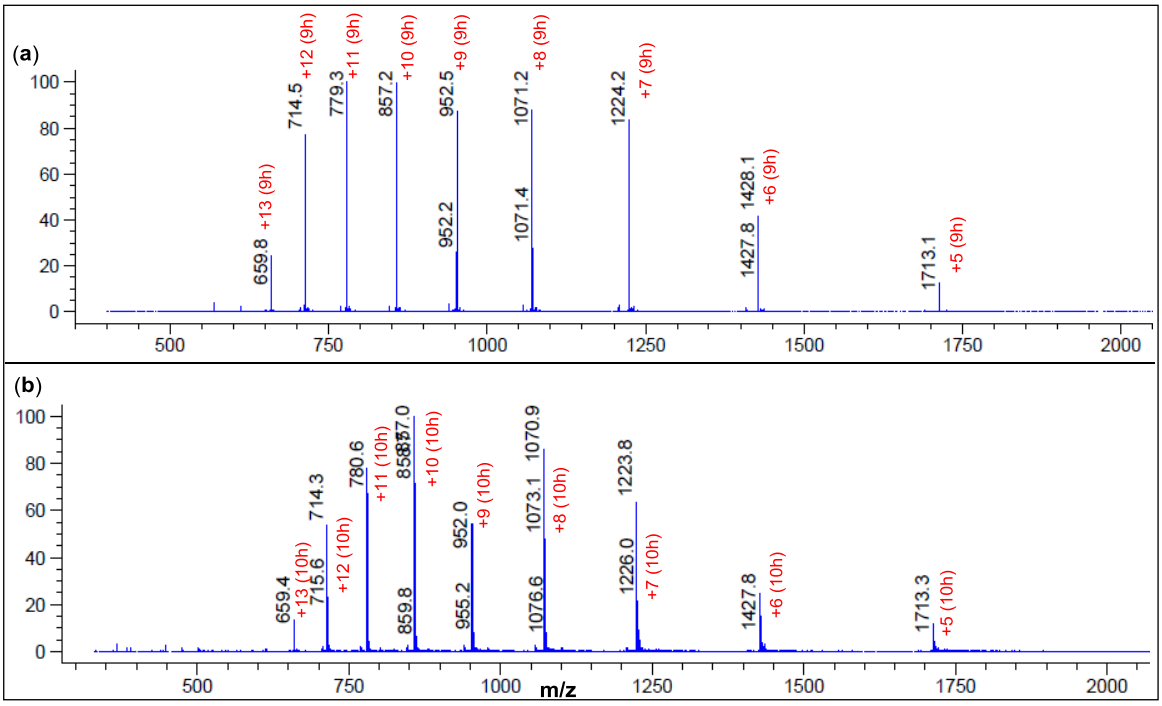
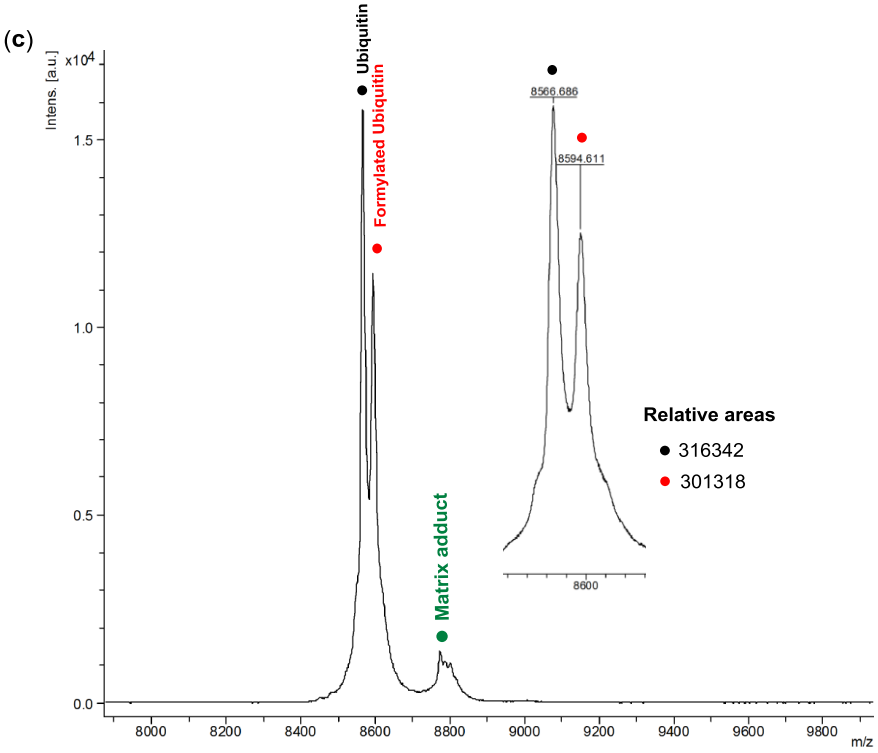


Figure S12. (a) MS spectra for Melittin **9g** (MW 2846 Da, 1 equiv.) and aldehyde (300 equiv.): mono-formylated product **10g**. (b) MS-MS spectrum of modified IKRKRQQ (20-26) from the digest of Melittin identifies K21 as the site of modification in mono-formylated Melittin **10g**.


(a)
659.8 +13 (9h)
714.5 +12 (9h)
779.3 +11 (9h)
857.2 +10 (9h)
952.5 +9 (9h)
952.2
1071.2 +8 (9h)
1071.4
1224.2 +7 (9h)
1427.8 1428.1 +6 (9h)
1713.1 +5 (9h)
(b)
659.4 +13 (10h)
715.6 +12 (10h)
714.3
780.6 +11 (10h)
857.0 +10 (10h)
859.8
952.0 +9 (10h)
955.2
1073.1 1070.9 +8 (10h)
1076.6
1223.8
1226.0 +7 (10h)
1427.8 +6 (10h)
1713.3 +5 (10h)
m/z


(c)
Intens. [a.u.]
x10^4
Ubiquitin
Formylated Ubiquitin
Matrix adduct
8566.686
8594.611
Relative areas
316342
301318
m/z

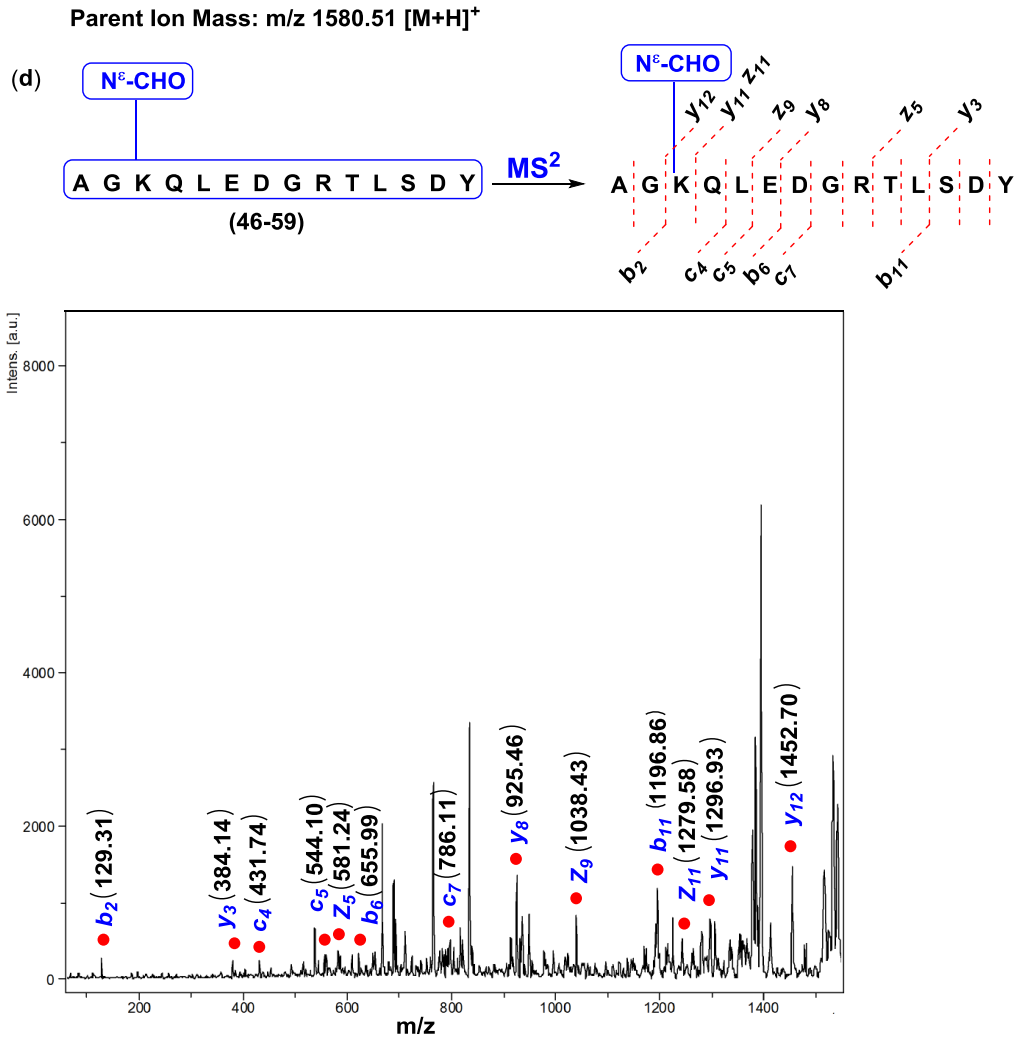
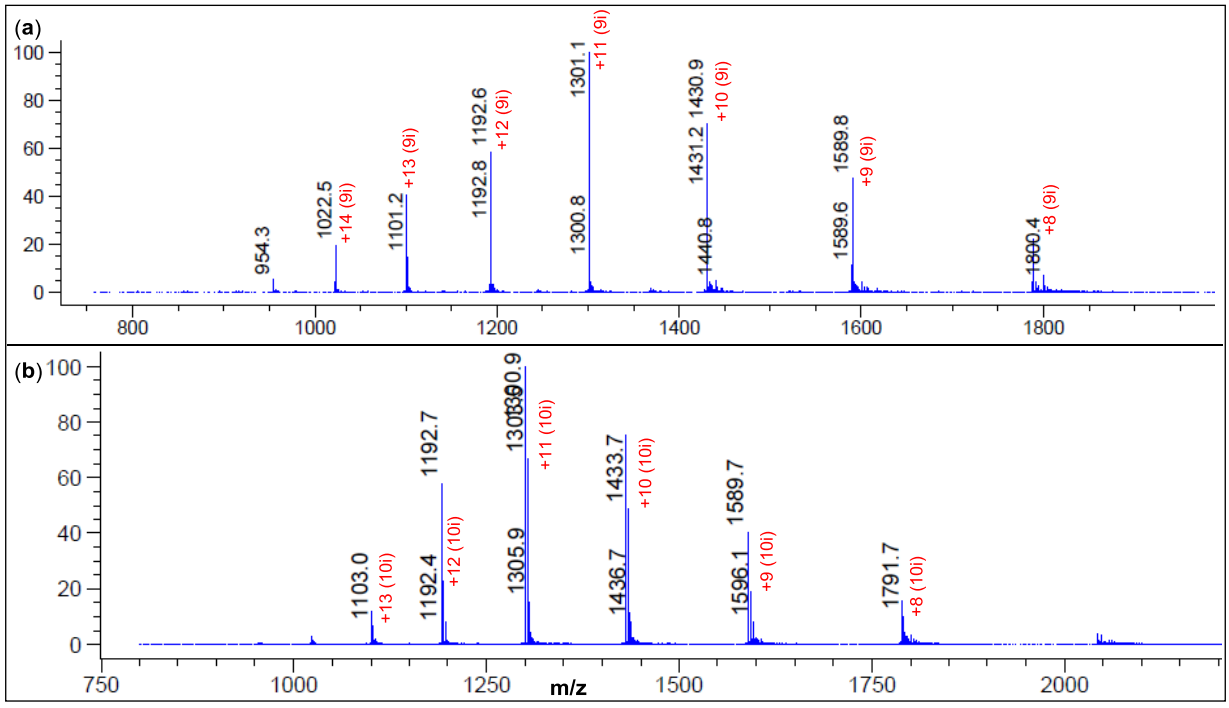
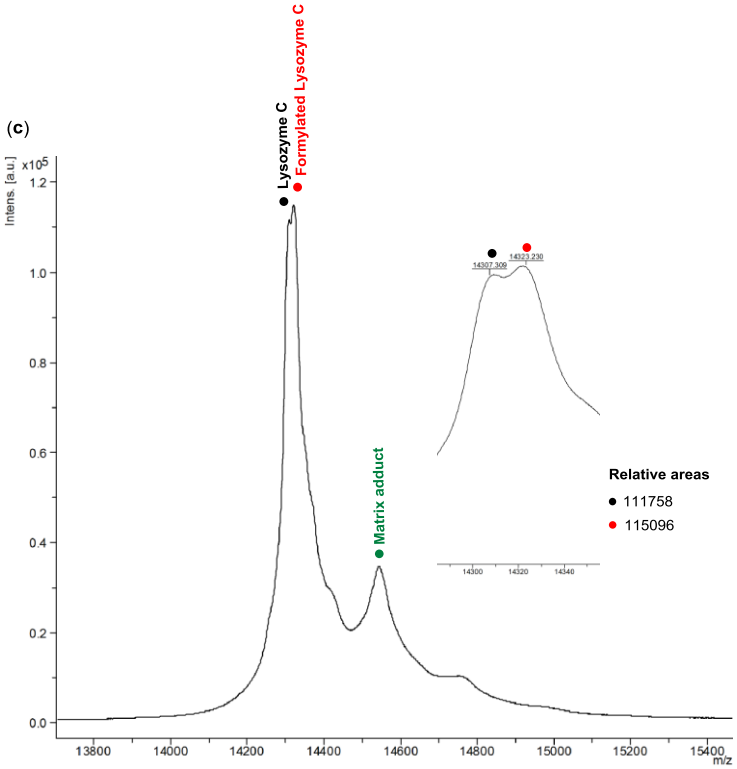


Figure S13. (a) ESI-MS spectra of native Ubiquitin **9h** (MW 8561 Da). (b) ESI-MS spectra of Ubiquitin (1 equiv.) and aldehyde (600 equiv.): mono-formylated product **10h**. (c) MALDI-MS spectra of Ubiquitin (1 equiv.) and aldehyde (600 equiv.): mono-formylated product **10h**. (d) MS-MS spectrum of AGKQLEDGRTLSDY (46-59). The site of modification is K48 in mono-formylated Ubiquitin **10h**.


(a)
954.3
1022.5
+14 (9i)
1101.2
+13 (9i)
1192.8
1192.6
+12 (9i)
1300.8
1301.1
+11 (9i)
1431.2
1430.9
+10 (9i)
1440.8
1589.6
1589.8
+9 (9i)
1800.4
+8 (9i)
(b)
1103.0
+13 (10i)
1192.4
1192.7
+12 (10i)
1305.9
+11 (10i)
1436.7
1433.7
+10 (10i)
1596.1
1589.7
+9 (10i)
1791.7
+8 (10i)
m/z


(c)
Intens. [a.u.]
Lysozyme C
Formylated Lysozyme C
Matrix adduct
Relative areas
111758
115096
m/z

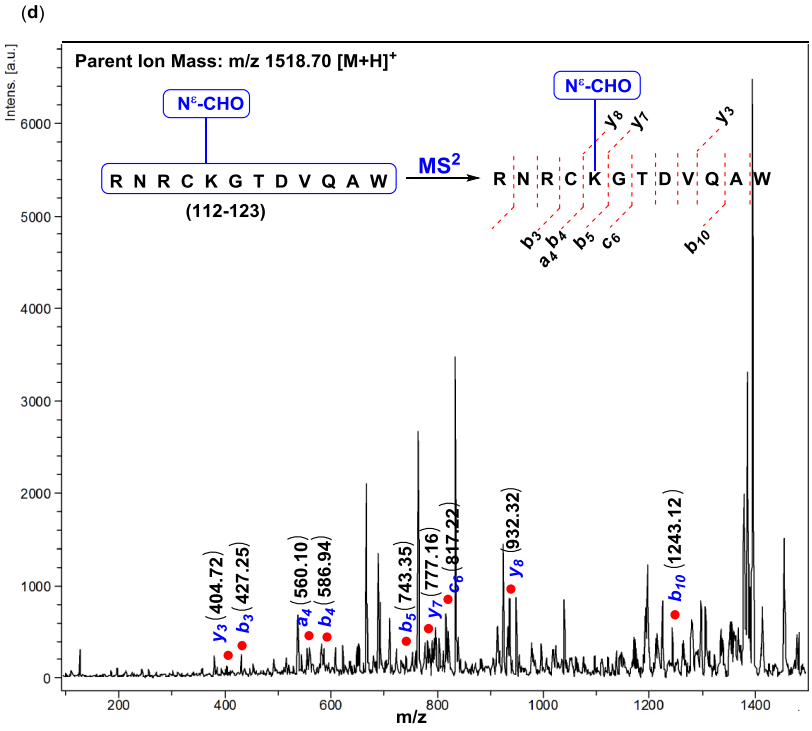
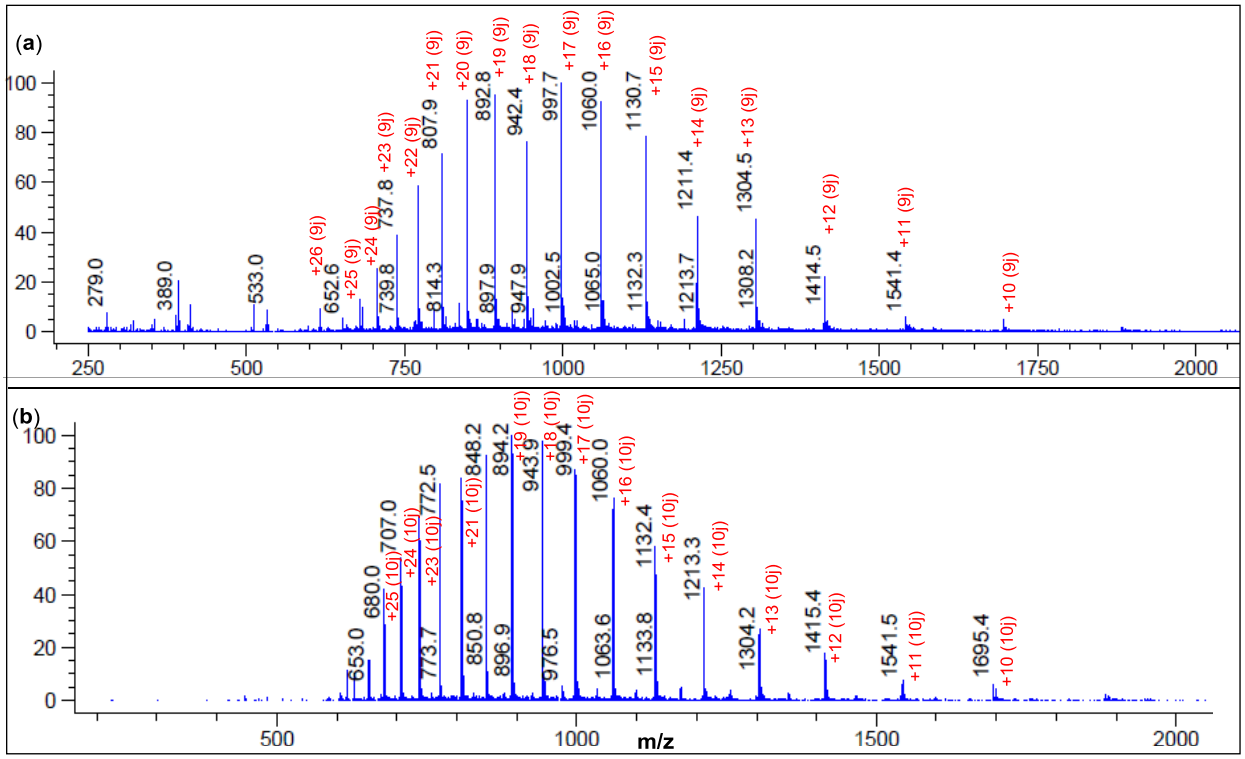
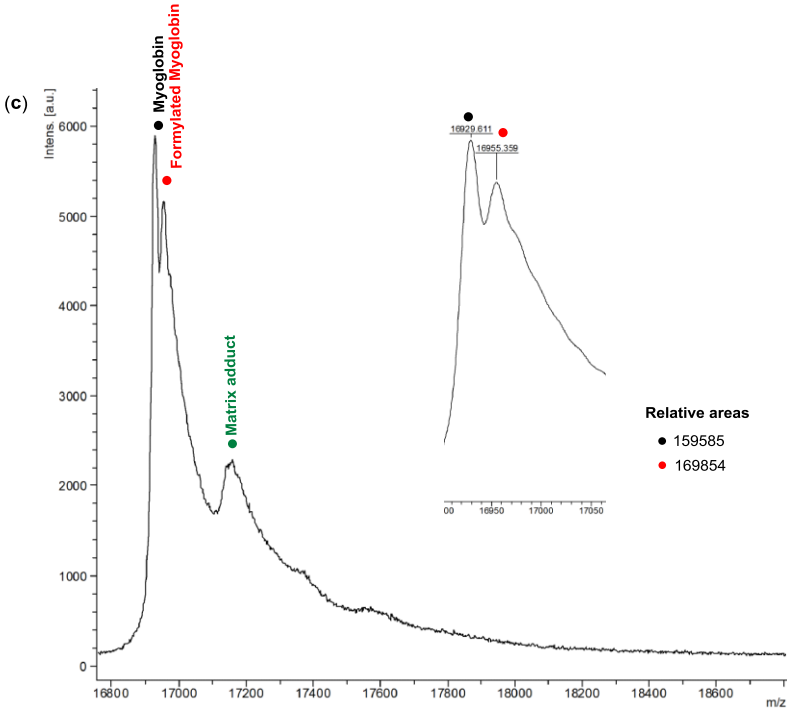


Figure S14. (a) ESI-MS spectra of native Lysozyme C **9i** (MW 14304 Da). (b) ESI-MS spectra of Lysozyme C (1 equiv.) and aldehyde (600 equiv.): mono-formylated product **10i**. (c) MALDI-MS spectra of Lysozyme C (1 equiv.) and aldehyde (600 equiv.): mono-formylated product **10i**. (d) MS-MS spectrum of RNRCKGTDVQAW (112-123). Site of modification is K116 in mono-formylated Lysozyme C **10i**.


(a)
279.0
389.0
533.0
652.6
739.8
737.8
814.3
807.9
892.8
897.9
942.4
947.9
997.7
1002.5
1060.0
1065.0
1130.7
1132.3
1211.4
1213.7
1304.5
1308.2
1414.5
1541.4
+26 (9j)
+25 (9j)
+24 (9j)
+23 (9j)
+22 (9j)
+21 (9j)
+20 (9j)
+19 (9j)
+18 (9j)
+17 (9j)
+16 (9j)
+15 (9j)
+14 (9j)
+13 (9j)
+12 (9j)
+11 (9j)
+10 (9j)
(b)
653.0
680.0
707.0
772.5
773.7
848.2
850.8
894.2
896.9
943.9
976.5
999.4
1060.0
1063.6
1132.4
1133.8
1213.3
1304.2
1415.4
1541.5
1695.4
+25 (10j)
+24 (10j)
+23 (10j)
+21 (10j)
+19 (10j)
+18 (10j)
+17 (10j)
+16 (10j)
+15 (10j)
+14 (10j)
+13 (10j)
+12 (10j)
+11 (10j)
+10 (10j)
m/z


(c)
Intens. [a.u.]
Myoglobin
Formylated Myoglobin
Matrix adduct
16929.611
16955.359
Relative areas
159585
169854
m/z

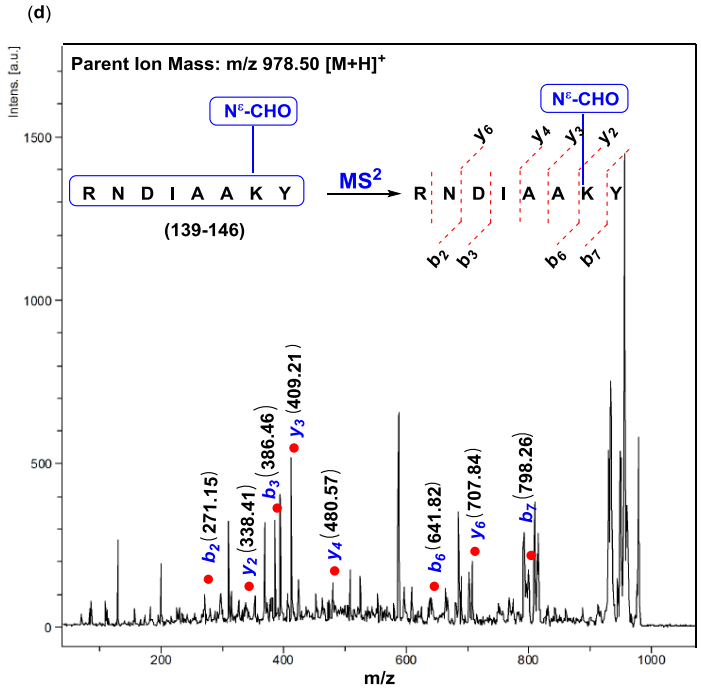
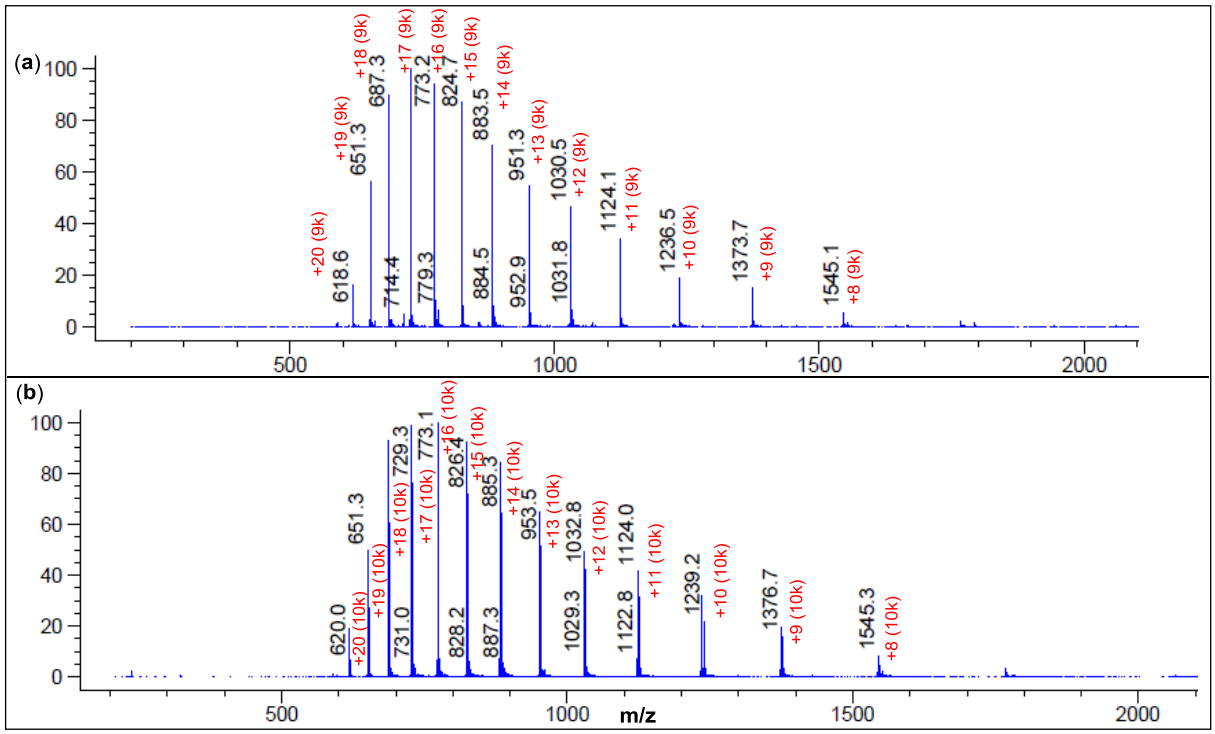


Figure S15. (a) ESI-MS spectra of native Myoglobin **9j** (MW 16951 Da). (b) ESI-MS spectra of Myoglobin (1 equiv.) and aldehyde (600 equiv.): mono-formylated product **10j**. (c) MALDI-MS spectra for Myoglobin **9j** (MW 16951 Da, 1 equiv.) and aldehyde (600 equiv.): mono-formylated product **10j**. (d) MS-MS spectrum of modified RNDIAAKY (139-146) from the digest of mono-formylated Myoglobin **10j** identifies K145 as the site of modification.



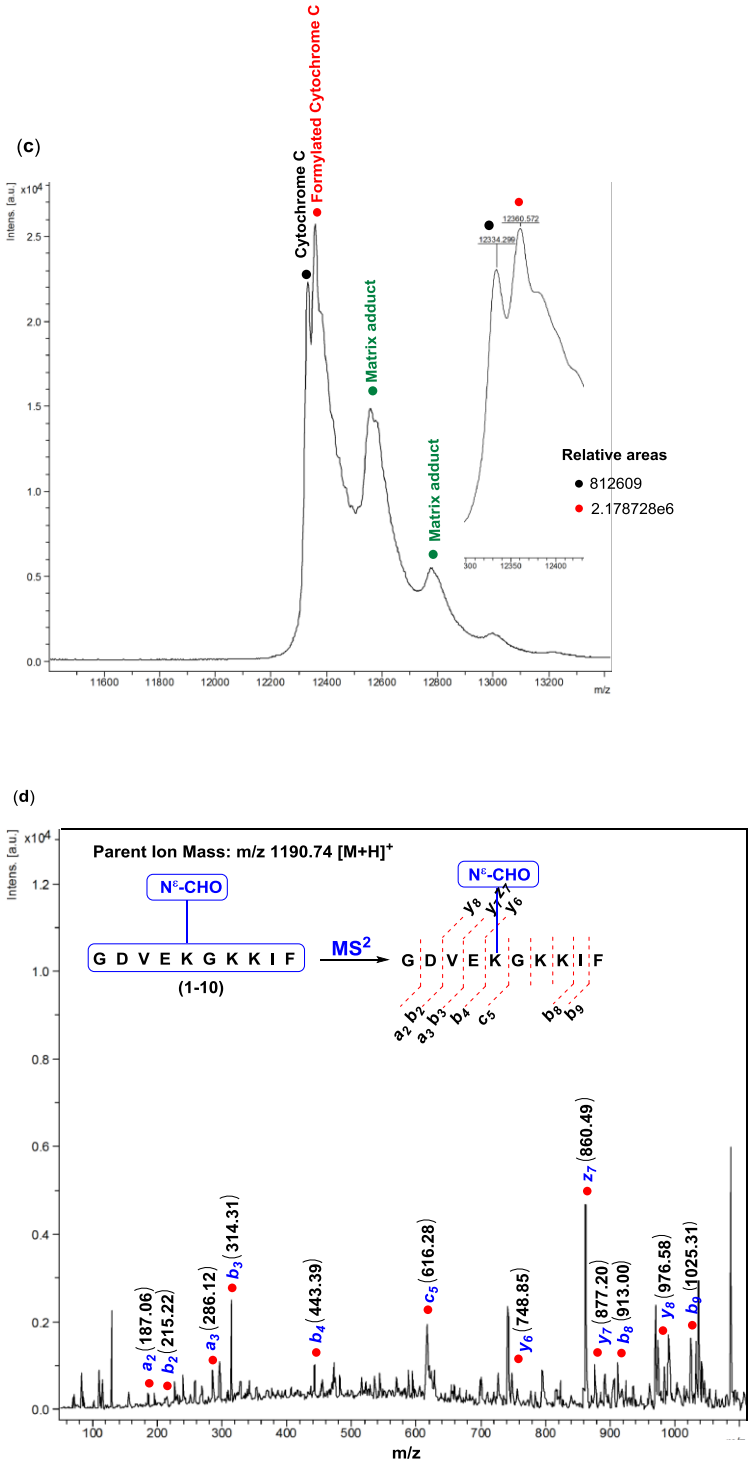


Figure S16. (a) ESI-MS spectra of native Cytochrome C **9k** (MW 12355 Da). (b) ESI-MS spectra of Cytochrome C (1 equiv.) and aldehyde (600 equiv.): mono-formylated product **10k**. (c) MALDI-MS spectra for Cytochrome C **9k** (MW 12355 Da, 1 equiv.) and aldehyde (600 equiv.): mono-formylated product **10k**. (d) MS-MS spectrum of modified GDVEKGKKIF (1-10) from the digest of mono-formylated Cytochrome C **10k** identifies K5 as the site of modification.

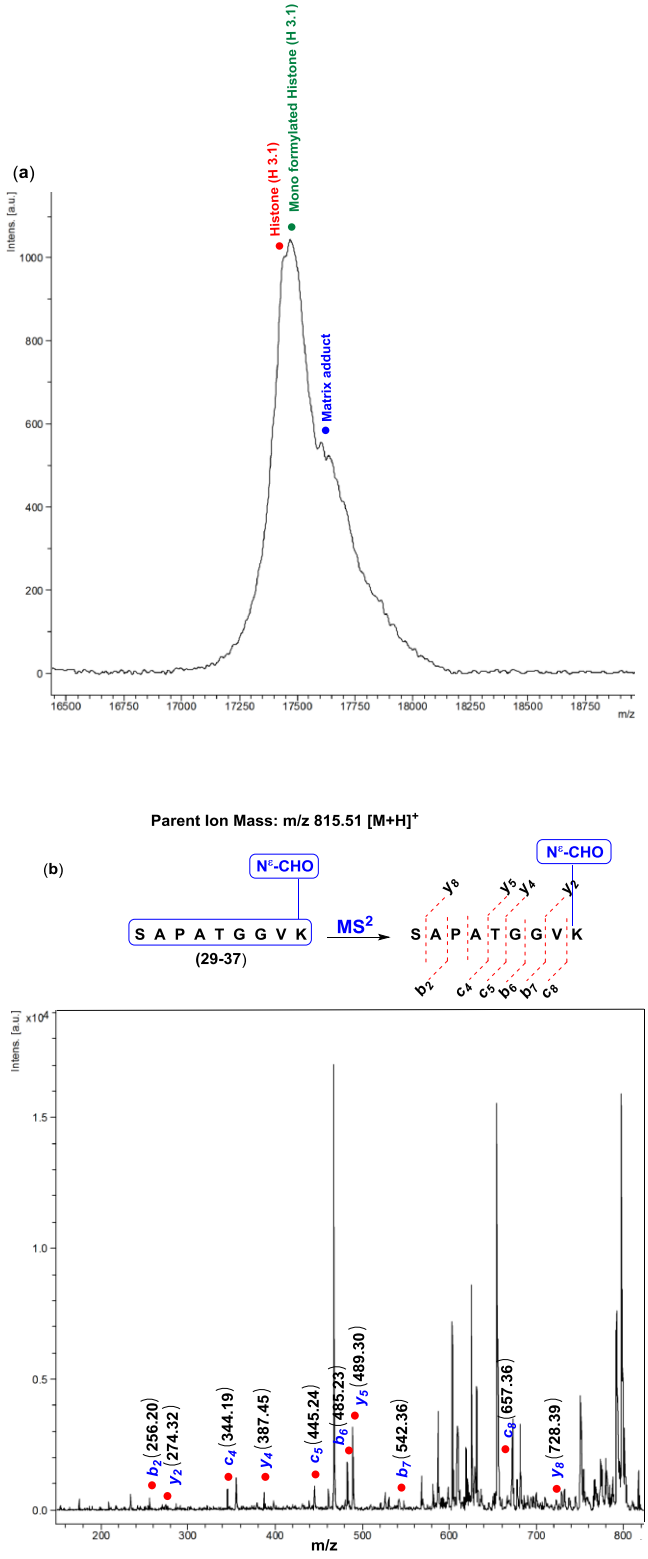


Figure S17. (a) MS spectra for Histone (H 3.1) **9l** (MW 17442 Da, 1 equiv.) and aldehyde (600 equiv.): mono-formylated product **10l**. (b) MS-MS spectrum of modified SAPATGGVK (29-37) from the digest of mono-formylated Histone (H 3.1) **10l** identifies K37 as the site of modification.

7.2. MS spectra for kinetic reactivity order of Lys residues in formylation of proteins using pre-synthesized formate ester 5a (*see Table S5*).

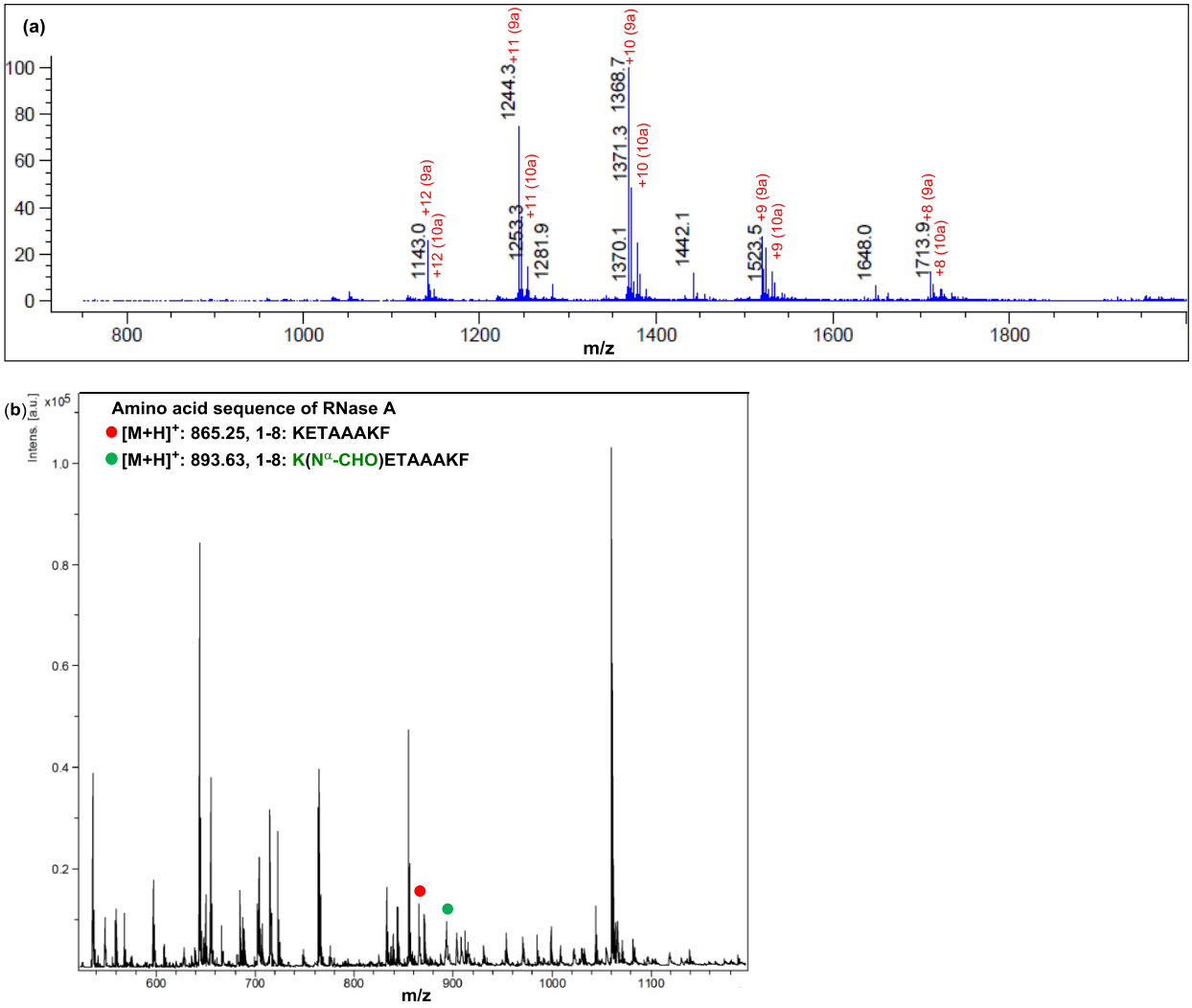
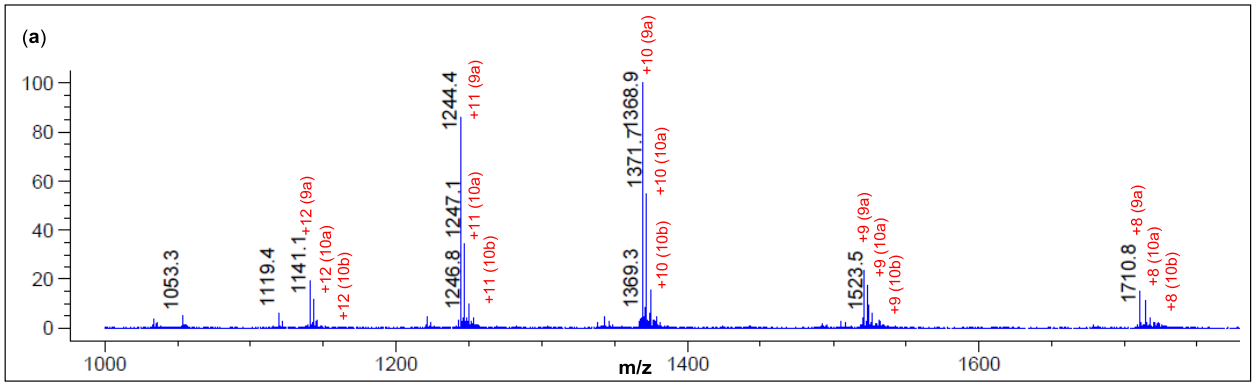


Figure S18. (a) ESI-MS spectra for RNase A **9a** (1 equiv.) vortexed with pre-synthesized formate ester **5a** (5 equiv., 1 h): mono-labeled formylated product (**10a**). (b) Peptide mapping: MS spectra of digest of mono-formylated RNase A (N^{α}-NH_2 labeling) (*refer Figure 4*).



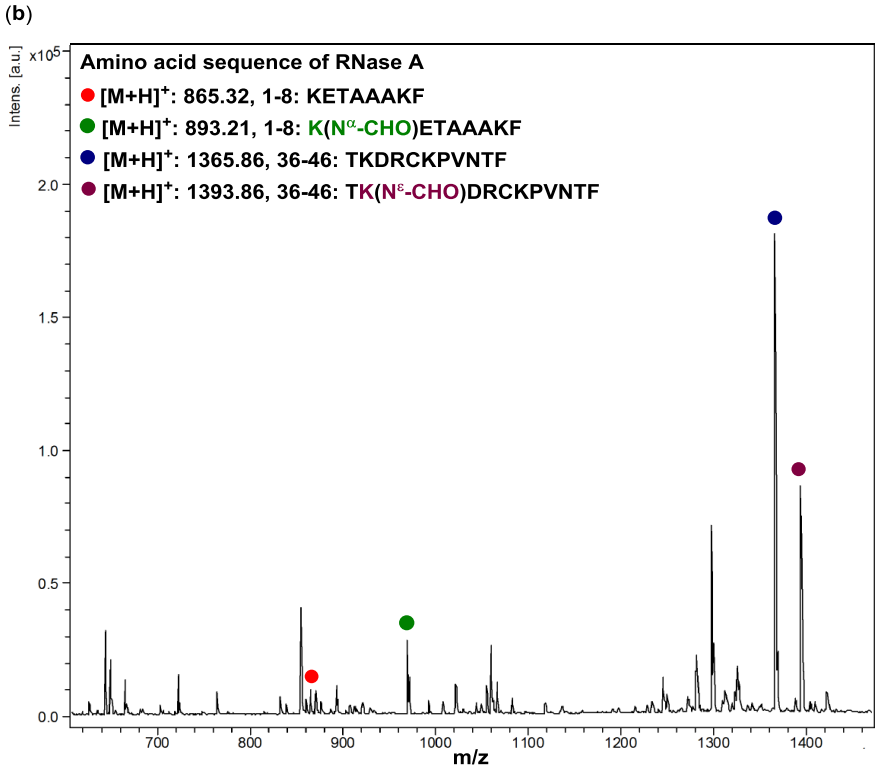


Figure S19. (a) MS spectra for RNase A (1 equiv.) vortexed with pre-synthesized formate ester **5a** (5 equiv., 3 h): mono-, bis-formylated product (**10a**, **10b**) (*refer Figure 4*). (b) Peptide mapping: MS spectra of digest of mono-, bis-formylated RNase A (*refer Figure 4*).

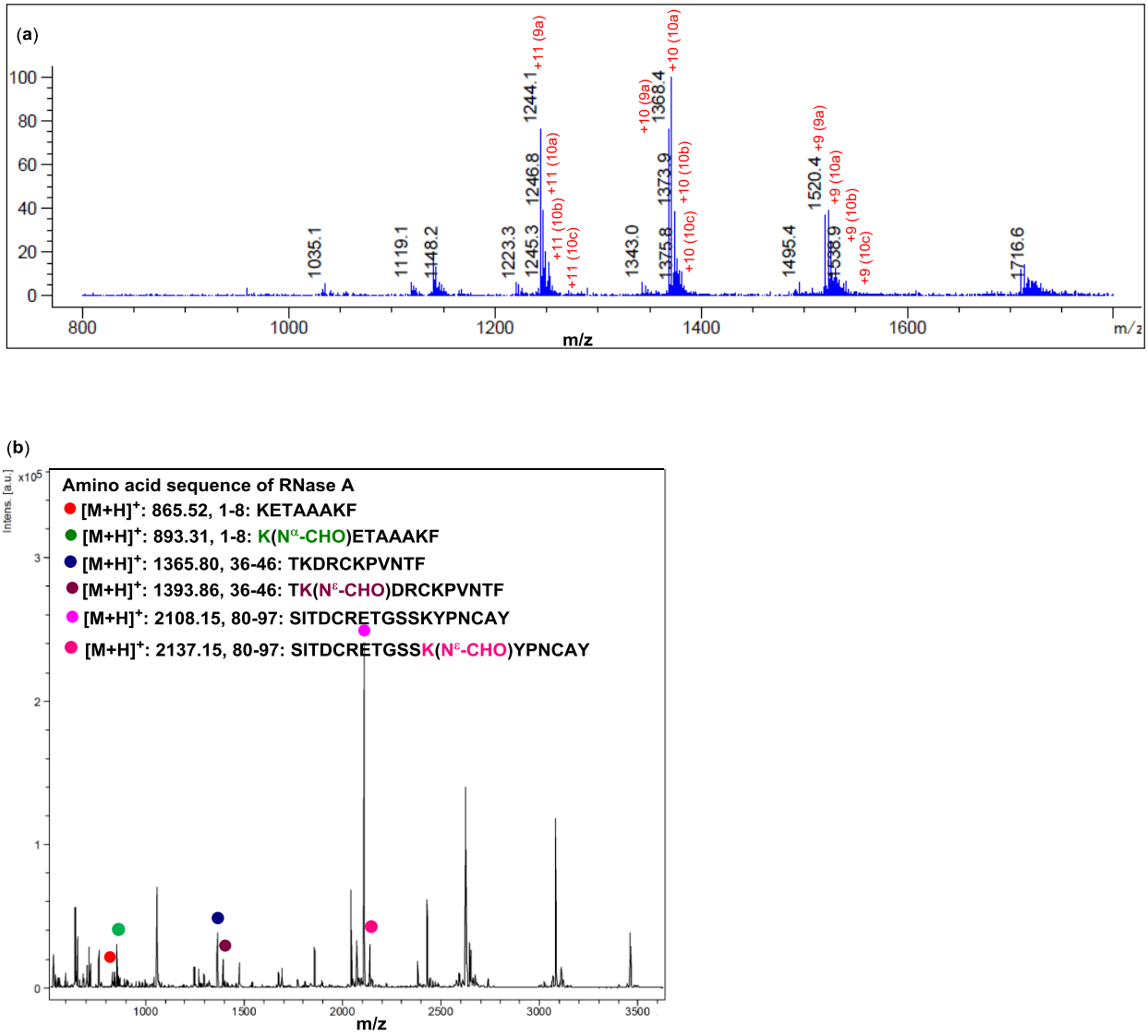


Figure S20. (a) MS spectra for RNase A (1 equiv.) vortexed with pre-synthesized formate ester **5a** (5 equiv., 5 h): mono-, bis- and tris-labeled formylated product (**10a**, **10b, 10c**) (*refer Figure 4*). (b) Peptide mapping: MS spectra of digest of mono-, bis- and tris-formylated RNase A **10a**-**10c** (*refer Figure 4*).

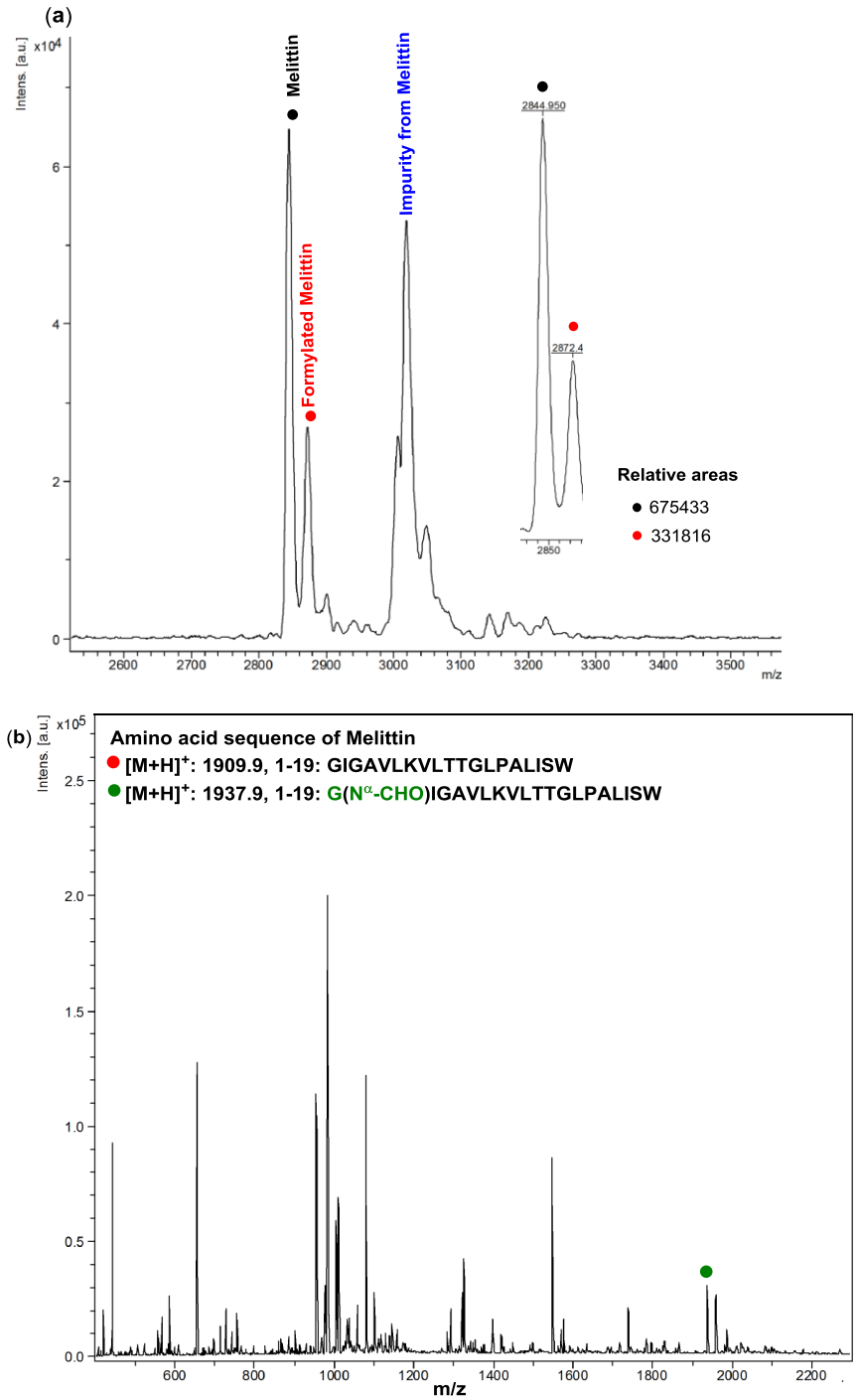


Figure S21. (a) MS spectra for Melittin (1 equiv.) vortexed with pre-synthesized formate ester **5a** (5 equiv., 1 h): mono-labeled formylated product (**S22**). (b) Peptide mapping: MS spectra of digest of mono-formylated Melittin.

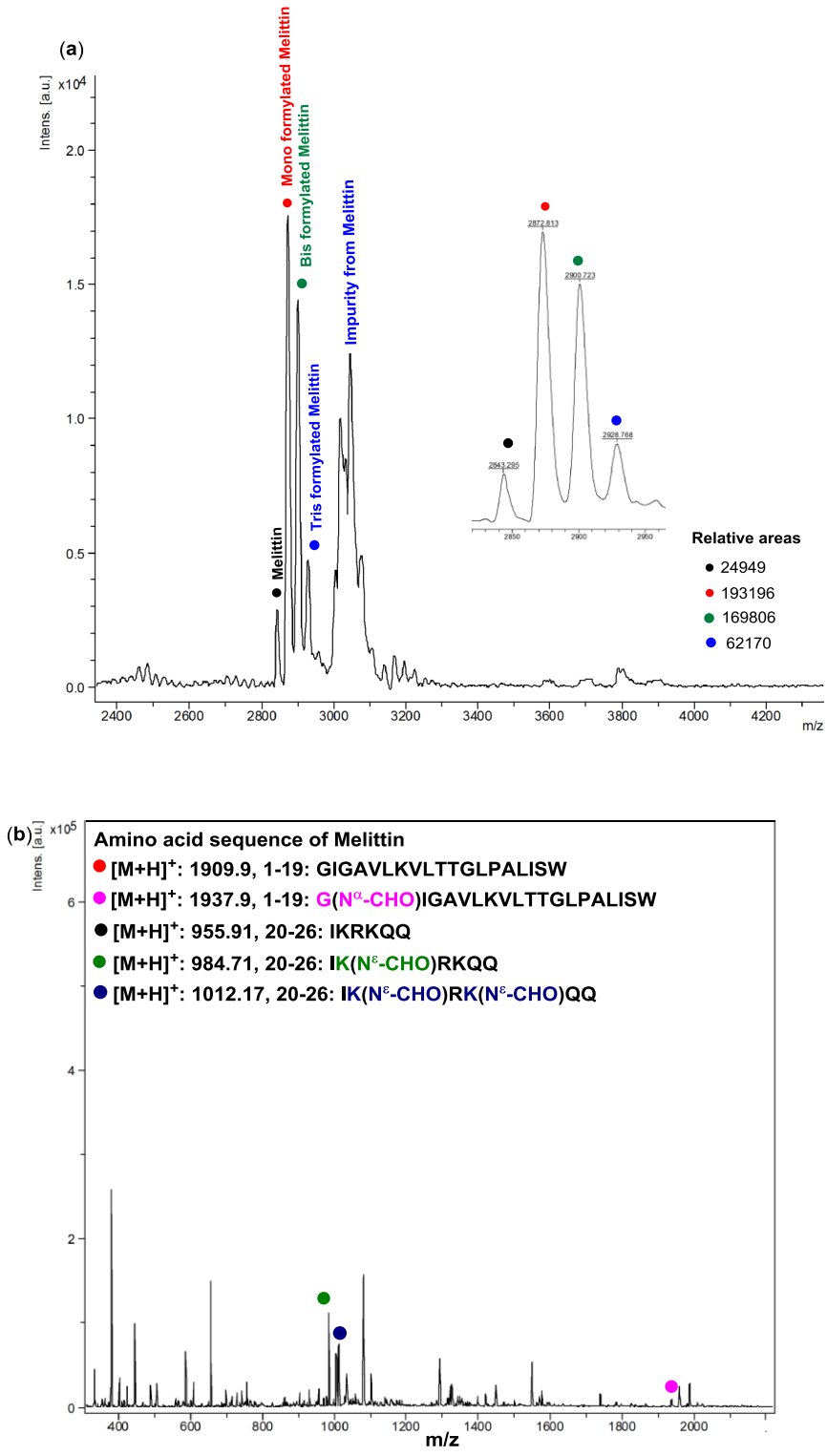


Figure S22. (a) MS spectra for Melittin (1 equiv.) vortexed with pre-synthesized formate ester **5a** (5 equiv., 5 h): mono-, bis- and tris-formylated Melittin (**S23**). (b) Peptide mapping: MS spectra of digest of mono-, bis- and tris-formylated Melittin.

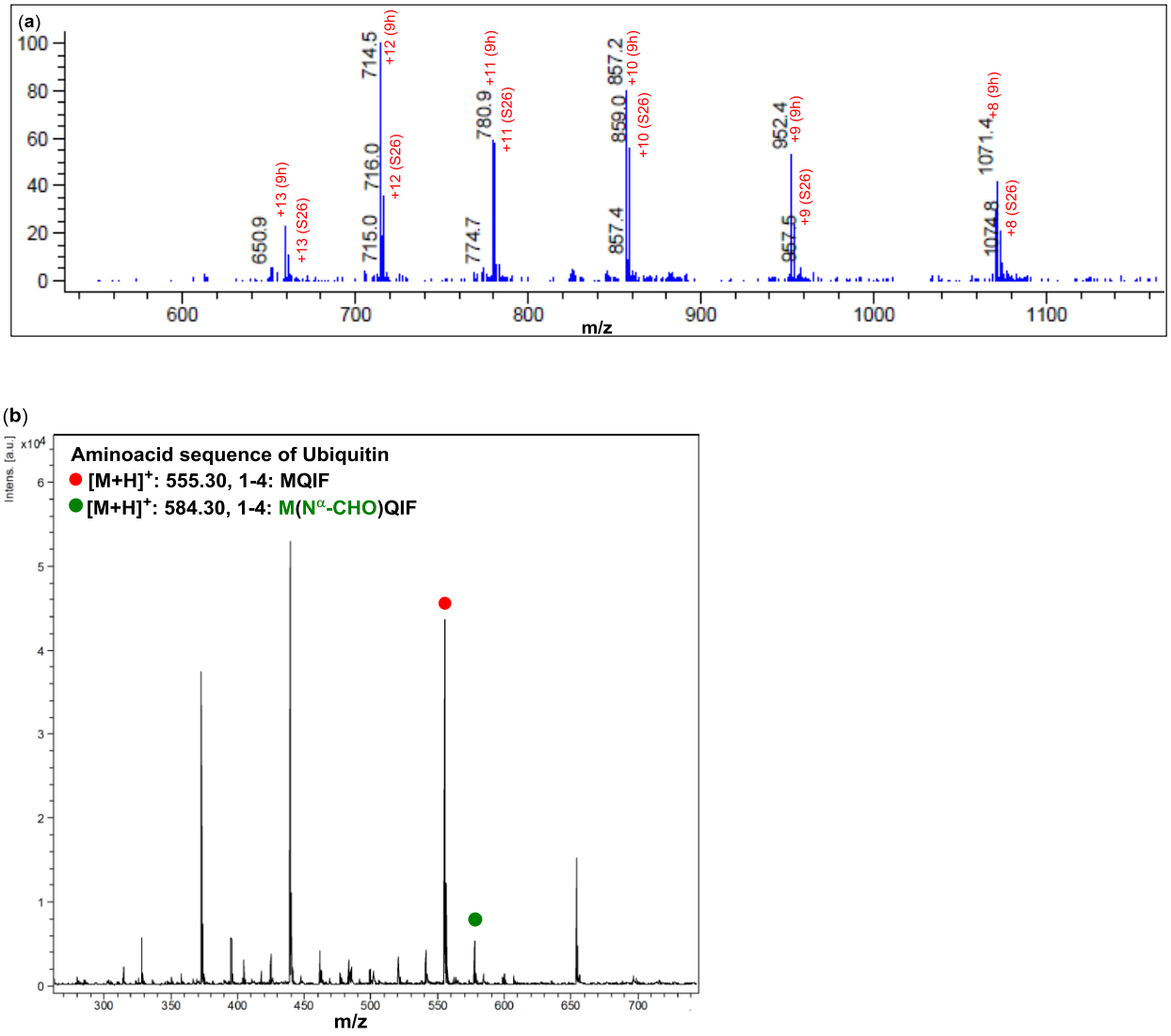


Figure S23. (a) MS spectra for Ubiquitin (1 equiv.) vortexed with pre-synthesized formate ester **5a** (3 equiv., 5 h) results in a mono-labeled formylated (**S26**) product. (b) Peptide mapping. MS spectra of digest of mono-formylated Ubiquitin.

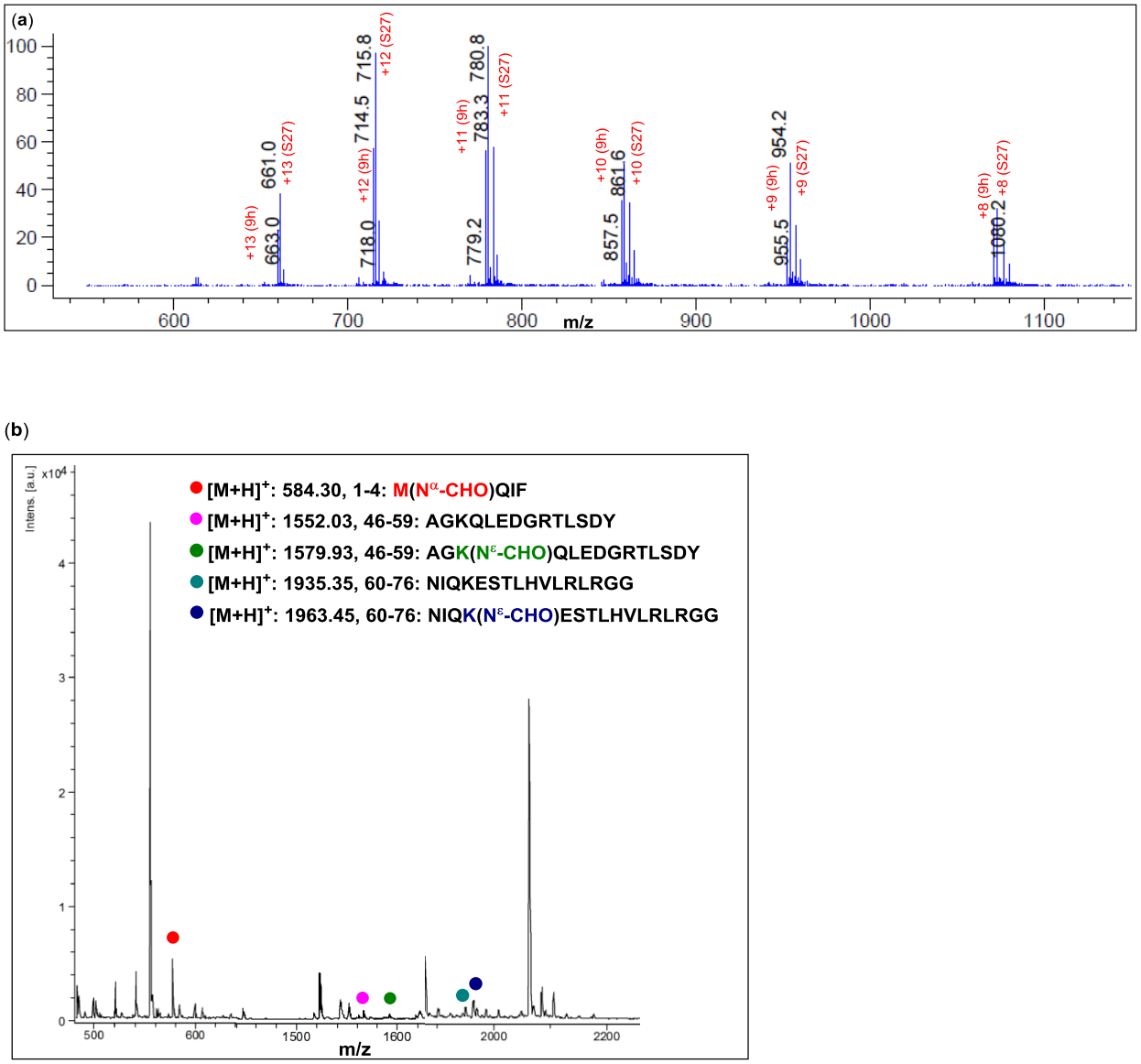


Figure S24. (a) MS spectra for Ubiquitin (1 equiv.) vortexed with pre-synthesized formate ester **5a** (5 equiv., 5 h) results in mono-, bis- and tris-formylated product (**S27**). (b) Peptide mapping. MS spectra of digest of mono-, bis- and tris-formylated Ubiquitin.

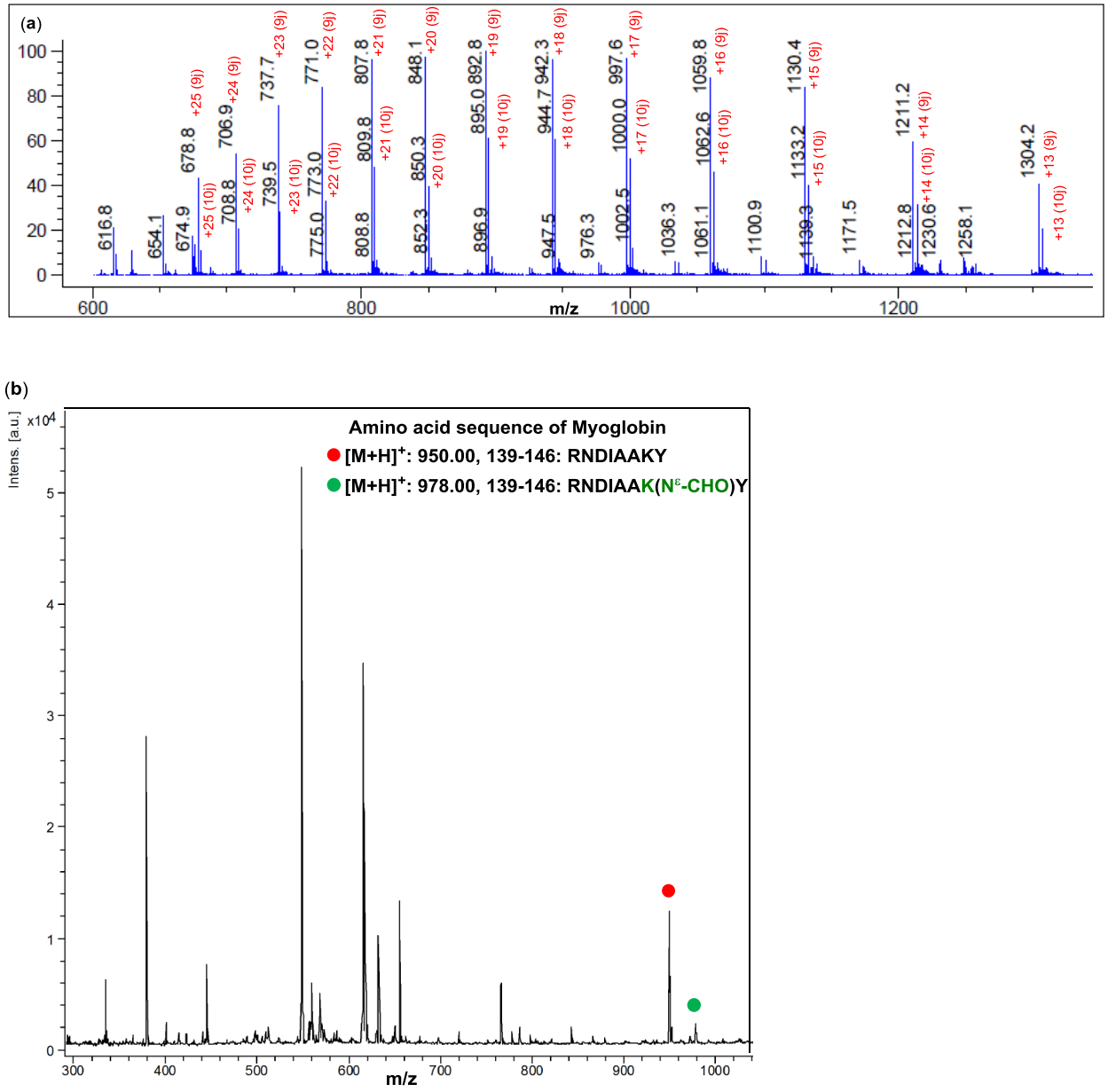
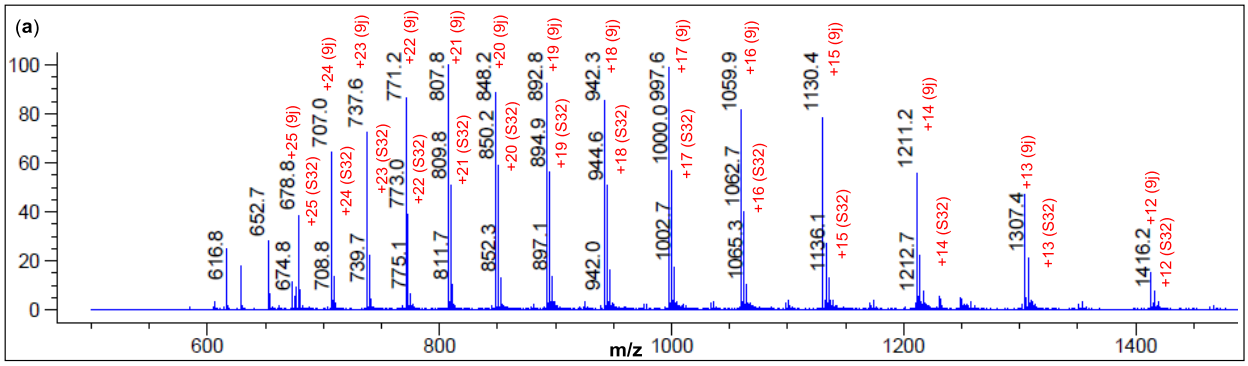


Figure S25. (a) MS spectra for Myoglobin (1 equiv.) vortexed with pre-synthesized formate ester **5a** (3 equiv., 5 h) results in a mono-labeled formylated product (**10j**). (b) Peptide mapping. MS spectra of digest of mono-formylated Myoglobin.



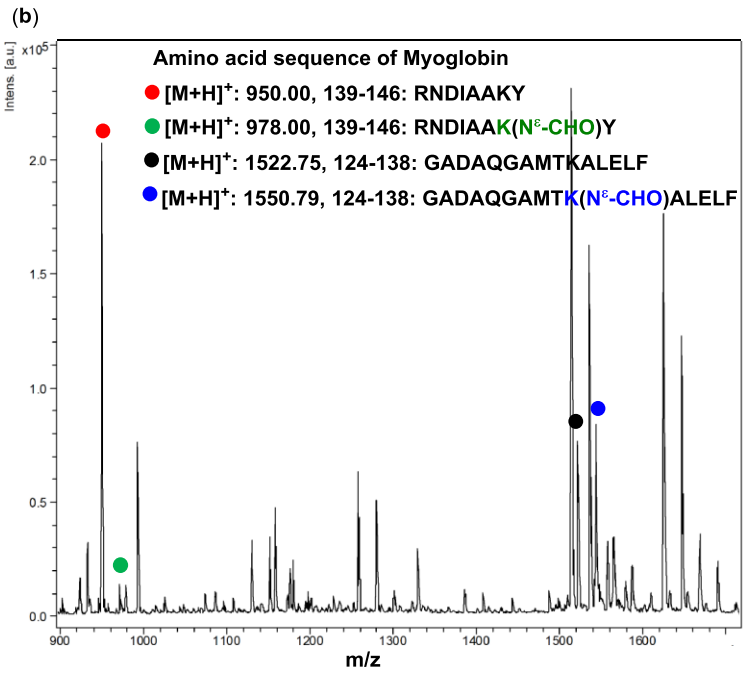


Figure S26. (a) MS spectra for Myoglobin (1 equiv.) vortexed with pre-synthesized formate ester **5a** (5 equiv., 5 h) results in mono and bis-formylated product.Peptide mapping. (b) MS spectra of digest of mono and bis-formylated Myoglobin.

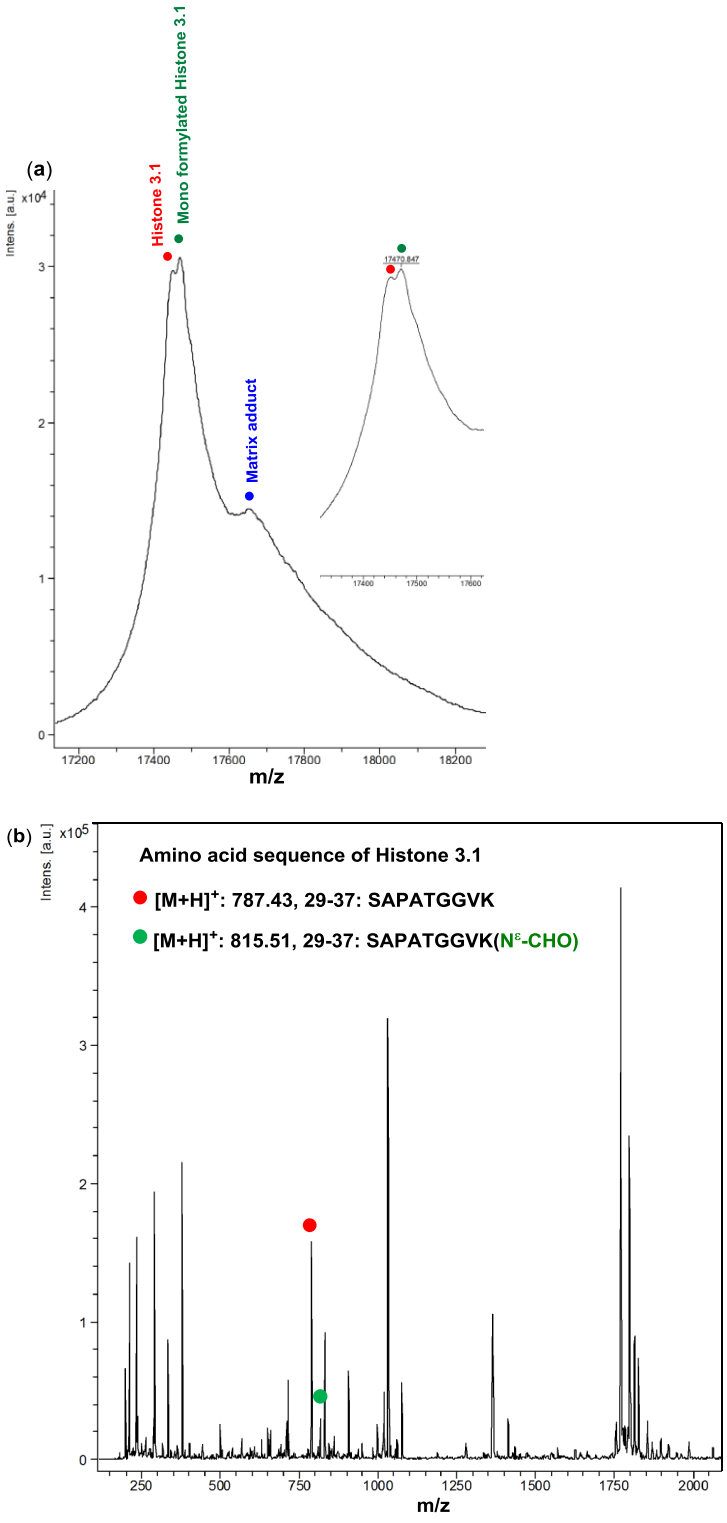


Figure S27. (a) MS spectra Histone (H 3.1) (1 equiv.) vortexed with pre-synthesized formate ester **5a** (5 equiv., 5 h): mono formylated Histone (H 3.1). (b) Peptide mapping: MS spectra of digest of mono formylated Histone (H 3.1).

7.3. MS spectra for interception of formylation (*refer Figure 6 and Figure S10*)

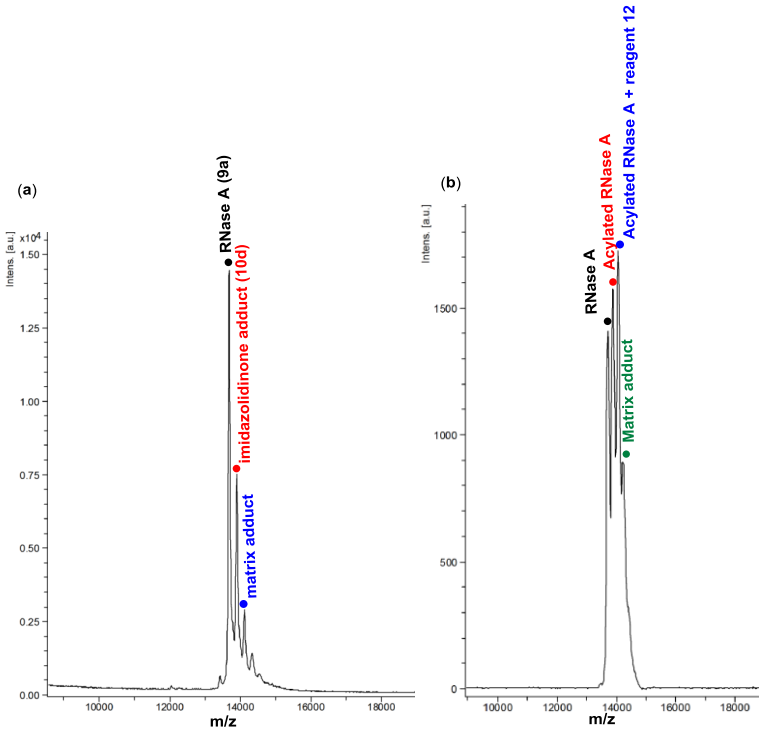


Figure S28. (a) N-terminus protection of RNase A through imidazolidinone formation using aldehyde **1a**. (b) MS spectra for acylated RNase A **10m** and labeled RNase A **13**. Interception of RNase A formylation by site-selective acylation and late-stage modification with 7-mercapto-4-methylcoumarin **12** results **13**.

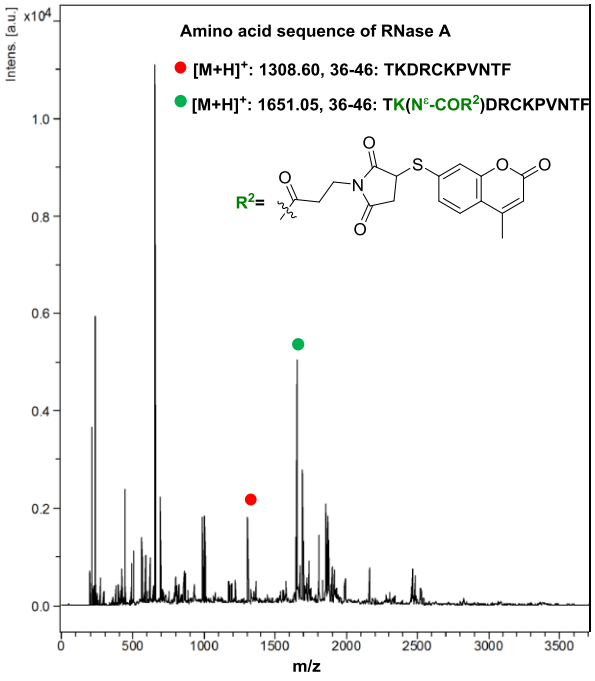


Figure S29. Peptide mapping. MS spectra of digest of monoacylated RNase A **13**.

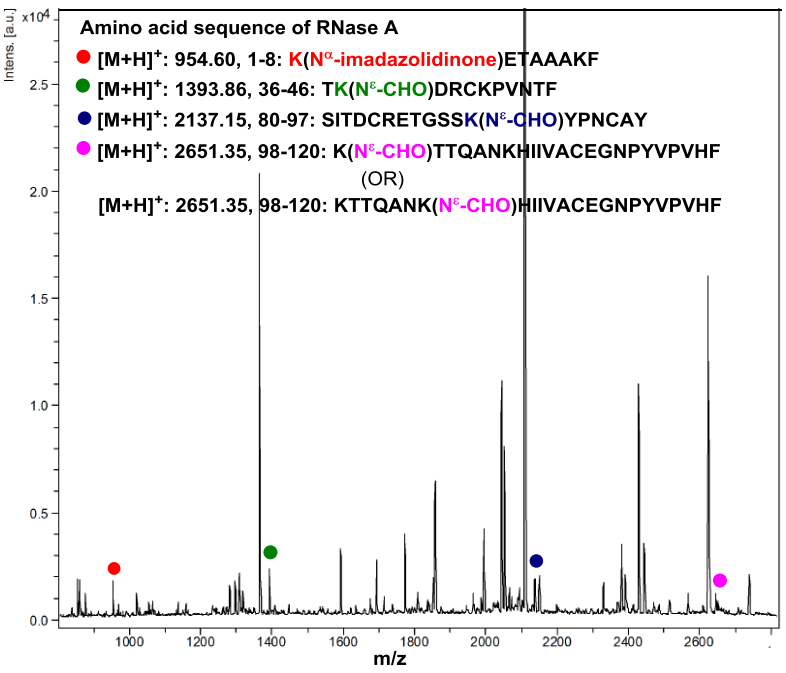


Figure S30. Peptide mapping of digest of mono-, bis- and tris-formylated RNase A (*refer Figure S8*).

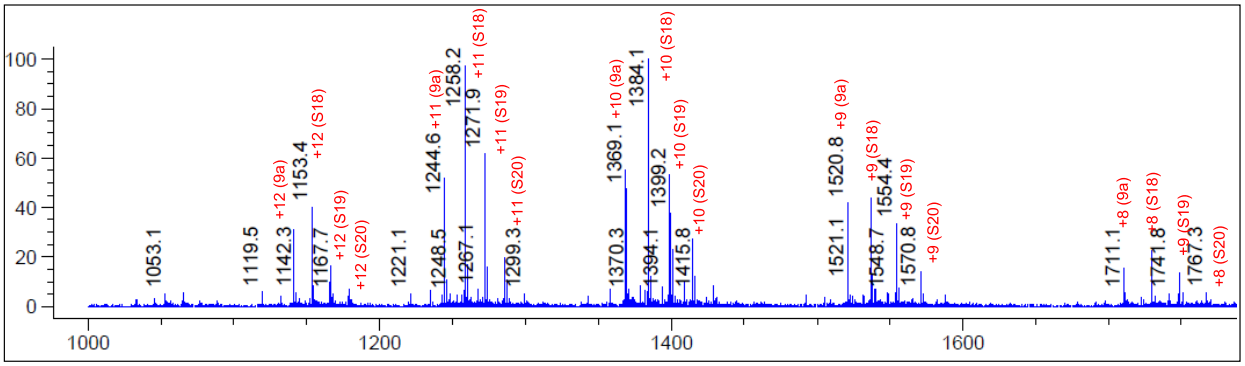


Figure S31. MS spectra for acylation of RNase A **9a** in the absence of aldehyde **1a** (*refer Figure S9*).

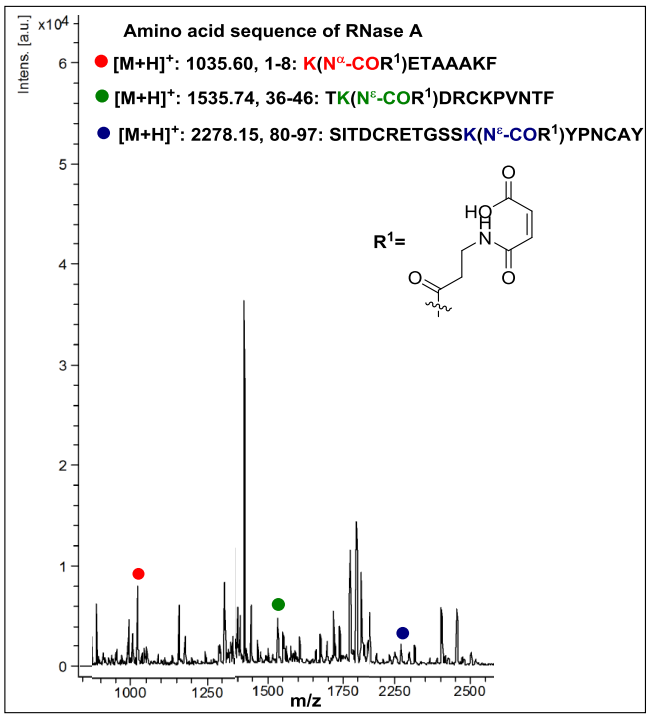


Figure S32. Peptide mapping of digest of mono-, bis- and tris- acylation of RNase A **9a** in the absence of aldehyde **1a** (*refer Figure S9*).

8. Formylation of proteins: additional control experiments

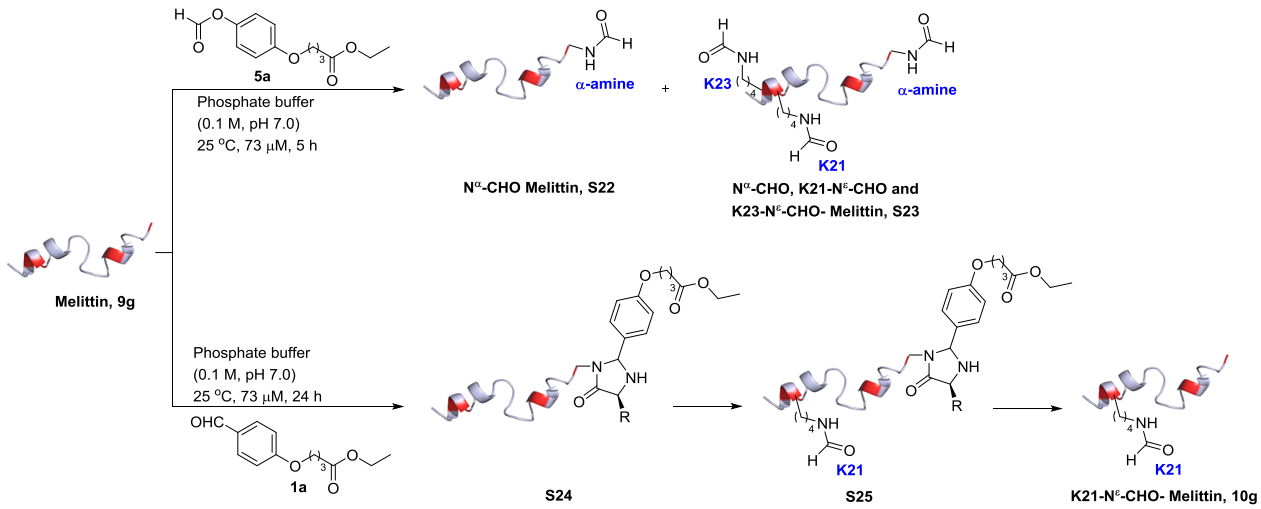


Figure S33. Melittin **9g** reacts with the formate **5a** and results in a heterogeneous mixture of **S22** and **S23**. The reaction with formylating pre-reagent aldehyde **1a** goes through site-selective and reversible generation of imidazolidinone **S25** to result in mono-formylated Melittin **10g**.

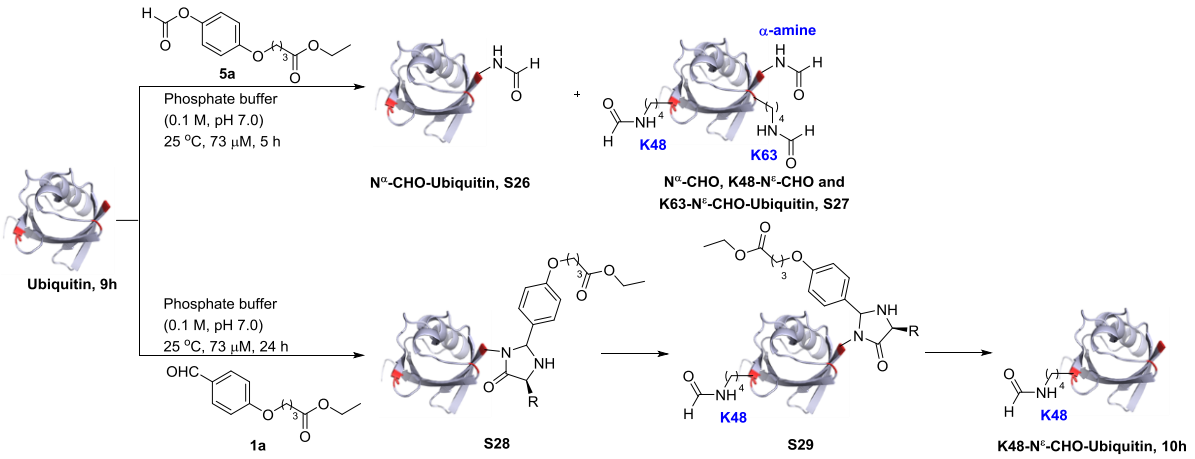


Figure S34. Ubiquitin **9h** reacts with the formate **5a** and results in heterogeneous mixture of **S26** and **S27**. The reaction with formylating pre-reagent aldehyde **1a** goes through site-selective and reversible generation of imidazolidinone **S29** to result in mono-formylated Ubiquitin **S10h**.

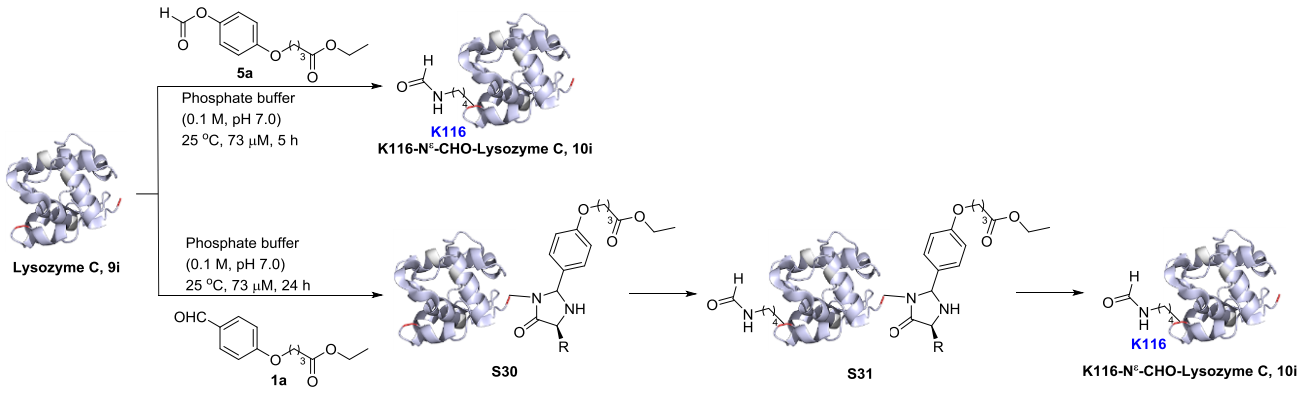


Figure S35. Lysozyme C **9i** reacts with the formate **5a** and results in monolabeled product **10i**. The reaction with formylating pre-reagent aldehyde **1a** goes through site-selective and reversible generation of imidazolidinone **S31** to result mono-formylated Lysozyme C **10i**.

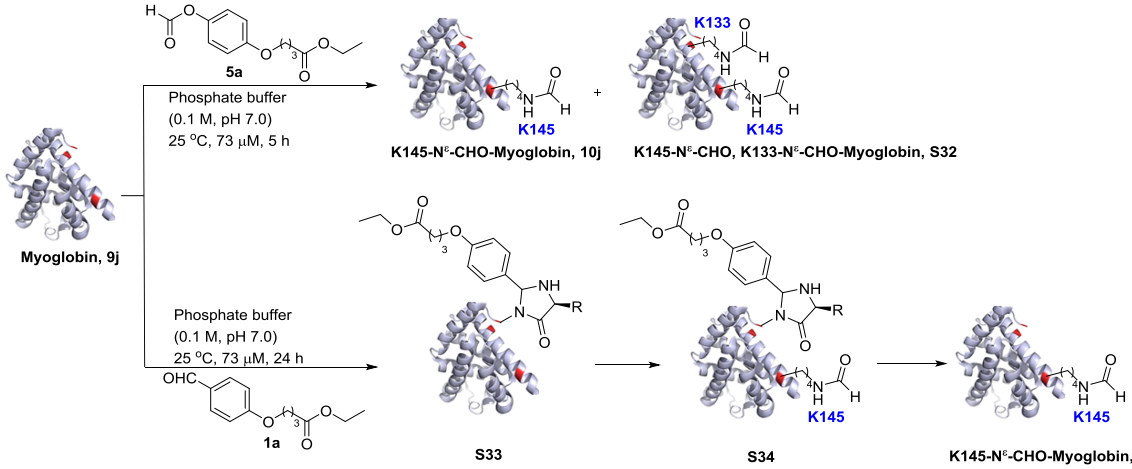


Figure S36. Myoglobin **9j** reacts with the formate **5a** and results in a heterogeneous mixture of **10j** and **S32**. The reaction with formylating pre-reagent aldehyde **1a** goes through site-selective and reversible generation of imidazolidinone **S34** to result in mono-formylated Myoglobin **10j**.

9. Enzymatic assay

9.1. Enzymatic assay of RNaseA[6]

RNase A activity before and after the labeling was checked by Ribonucleic acid (RNA) hydrolysis at 300 nm (A_{300}) using quartz cuvette (path length, 1 cm at 25 °C). Sodium acetate buffer (pH 5.0, 0.1 M) was prepared by using the Millipore Grade I water (pH was adjusted with 2 M acetic acid). Ribonucleic acid [RNA, 0.1% (w/v), 1 mg/ml] solution was prepared in sodium acetate buffer. Change in absorption at 300 nm of the RNA was performed using RNA solution and blank. Freshly prepared RNase A and labeled RNase A solutions (10 μg/1 ml, in Millipore Grade I water) were used for the assay. RNA (500 μl) and

RNase A (500 µl) solutions were mixed by inversion and immediately recorded at 300 nm for 10 min. Similarly, absorption of a mixed solution of RNA and labeled RNase A were performed. The enzymatic activity of RNase A remains unperturbed after the chemical modification.

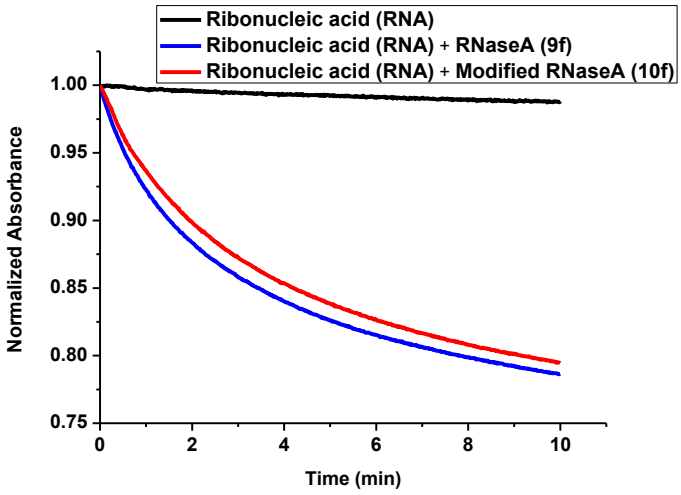


Figure S37. Normalized UV-Vis spectra. Comparison of enzymatic activity between native RNase A (**9f**) and formylated RNase A (**10f**). The enzymatic activity of RNase A remains unperturbed after the chemical modification. Absorbance was normalized to its values by dividing its absorbance maxima (OriginPro 8).

9.2. Enzymatic assay of Lysozyme C[7]

Lysozyme C activity before and after the labeling was checked by Micrococcus lysodeikticus lysis at 450 nm (A_{450}) using quartz cuvette (path length, 1 cm at 25 ^{o}C). Potassium phosphate buffer (pH 6.2, 0.1 M) was prepared by dissolving potassium phosphate, monobasic in the Millipore Grade I water. The pH was adjusted to 6.2 at 25 °C using 1 M potassium hydroxide (KOH) solution. Micrococcus lysodeikticus suspension [0.01% (w/v)] was prepared in phosphate buffer. The change in absorbance at 450 nm of this suspension versus a buffer blank was in agreement with the literature (0.6-0.7) after adjustment using the appropriate amount of buffer. Freshly prepared Lysozyme C and labeled Lysozyme C solutions (10 µg/1 ml, in phosphate buffer) were used for the assay. 100 µl of Lysozyme C solution and Micrococcus lysodeikticus suspension (1 ml) were mixed by inversion. The sample was immediately used for recording the absorbance at 450 nm. The protocol was repeated for measuring the absorption of Micrococcus lysodeikticus and labeled Lysozyme C mixture. The enzymatic activity of Lysozyme C remains unperturbed after the chemical modification.

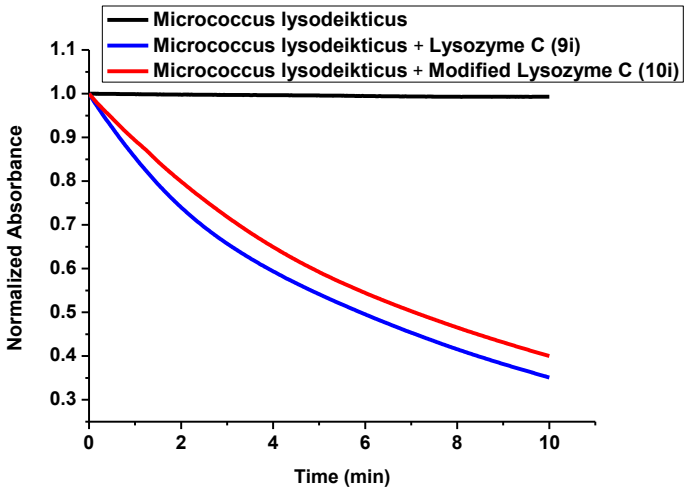
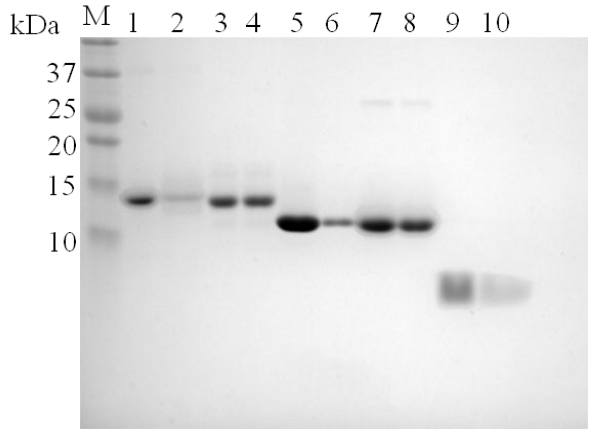


Figure S38. Comparison of Lysozyme C activity before and after reacting with aldehyde **1a.** The bioactivity was examined by Lysozyme assay purchased from Sigma-Aldrich.

10. SDS-PAGE



Coomassie

Figure S39. SDS-PAGE (15%) analysis shown with Coomassie staining. (a) Lane M= molecular weight marker; lane 1= Myoglobin (**9j**); lane 2 = formylated Myoglobin (**10j**); lane 3 = Lysozyme C (**9i**); lane 4 = formylated Lysozyme C (**10i**); lane 5 = RNase A (**9f**); lane 6 = formylated RNase A (**10f**); lane 7 = Cytochrome C (**9k**); lane 8 = formylated Cytochrome C (**10k**); lane 9 = Ubiquitin (**9h**); lane 10 = formylated Ubiquitin (**10h**).

11. Acknowledgements

L.P., S.R.A. and M.C. are the recipients of a research fellowship from CSIR and UGC, India. V.R. is a Ramanujan Fellow (SERB, India). We thank Dr. S. Mukherjee for helpful discussions and Dr. R. S. Tomar and Ms. S. Chauhan for histone (H 3.1) sample. The authors are thankful to SERB, DAE and DBT, India for research grants.

12. References

1 For expression and purification of recombinant human histones, see: Y. Tanaka, M. Tawaramoto-Sasanuma, S. Kawaguchi, T. Ohta, K. Yoda, H. Kurumizaka, S. Yokoyama *Methods*., 2004, **33**, 3.
2 S. P. Velagapudi, S. J. Seedhouse, J. French and M. D. Disney, *J. Am. Chem. Soc.,* 2011, **133**, 10111.
3 J. Han, L. Sun, Y. Chu, Z. Li, D. Huang, X. Zhu, H. Qian and W. Huang, *J. Med. Chem.,* 2013, **56**, 9955.
4 Z. Zhou, X. Wu, A. Kresak, M. Griswold and Z- R. Lu, *Biomaterials.*, 2013, **34**, 7683.
5 M. Kinter and N. E. Sherman, *Wiley Interscience*, 2000.
6 Link for the RNase A enzymatic assay by using the Ribonucleic acid http://www.sigmaaldrich.com/technical-documents/protocols/biology/enzymatic-assay-of ribonuclease-a.html
7 Link for the Lysozyme C enzymatic assay by using the Micrococcus lysodeikticus http://www.sigmaaldrich.com/technical-documents/protocols/biology/enzymatic-assay-of-lysozyme.html

Additional relevant references

8 J. I. MacDonald, H. K. Munch, T. Moore and M. B. Francis, *Nat. Chem. Biol*., 2015, **11**, 326
9 C. T. Walsh, S. Garneau-Tsodikova and Jr. G. Gatto, *Angew. Chem. Int. Ed.,* 2005, **44**, 7342.
10 M. Fernandez-Suarez, H. Baruah, L. Martínez-Hernández, K. T. Xie, J. M. Baskin, C. R. Bertozzi and A.Y Ting, *Nature Biotechnol.*, 2007, **25**, 1483.
11 N. A. Afagh and A. K. Yudin, *Angew. Chem. Int. Ed.,* 2010, **49**, 262.
12 I. S. Carrico, *Chem. Soc. Rev.*, 2008, **37**, 1423.
13 R. T. Dean, S. Fu, R. Stocker and M. J. Davies, *Biochem. J.*, 1997, **324**, 1.
14 C. Marteau, F. Ruyffelaere, J.-M. Aubry, C. Penverne, D. Favier and V. Nardello-Rataj, *Tetrahedron*, 2013, **69**, 2268.
15 A. Chanda and V. V. Fokin, *Chem. Rev*., 2009, **109**, 725.
16 J. Baeza, M. J. Smallegan and J. M. Denu, *ACS Chem. Biol.*, 2015, **10**, 122.
15 V. Gautier, A. J. Boumeester, P. Lossl and A. J. R. Heck, *Proteomics*, 2015, **15**, 2756.
16 D. Schumacher and C. P. R. Hackenberger, *Curr. Opin. Chem. Biol*., 2014, **22**, 62.
17 G. Walsh and R. Jefferis, *Nat. Biotechnol*., 2006, **24**, 1241.
18 S. Xiao, V. Patsalo, B. Shan, Y. Bi, D. F. Green and D. P. Raleigh, *Proc. Natl. Acad. Sci. U. S. A*., 2013, **110**, 11337.
19 E. M. Sletten and C. R. Bertozzi, *Angew. Chem., Int. Ed*., 2009, **48**, 6974.
20 M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner and H. L. Ploegh, *Nat. Chem. Biol*., 2007, **3**, 707.
21 C. P. R. Hackenberger and D. Schwarzer, *Angew. Chem., Int. Ed*., 2008, **47**, 10030.
22 J. R. Wisniewski, A. Zougman and M. Mann, *Nucleic Acids Res.*, 2008, **36**, 570.